

A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-709 <TIGR>
 A:Cross-references: GB:U32710; GB:LA2023; NID:G1573200; PID:G1573211; TIGR:HI0248

alignment_scores:

Quality: 2471.50 Length: 700.
 Ratio: 4.092 Gaps: 7
 Percent Similarity: 86.286 Percent Identity: 66.857

alignment_block:

US-09-303-518d-651 x C64057 ..

Align seg 1/1 to: C64057 from: 1 to: 709

```

79 TACTTGGCATATGCTGCTGCTGGCATCTTCCCAAGCTTGGGCGG 128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
10 PheLeuThrAlaCysValSerLeuGlyIleAlaSerGlnAlaIleAlaIle 26
129 ACACATTTATTTCCGCATCACTACCAATATCTATCGGACTTTGGCGAAA 178
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
26 yHstHrTYrPheGlyIleAspTYrGlnTYrTYrArgAspPheAlaGlu 43
179 ATAAAGCAAGTTTGAGTCGGGCGGAAAGATATTGAGGTTTACACAAA 228
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
43 snlysgIlySpherThrValGlyAlaIlyAsnIleGlyValIlyAsnIly 59
229 AAAGGAGCTTGTGGCAAAATCAATGACAAAAGCCGATGATGATT 278
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
60 GluGlyGlnLeuValGlyThrSerMetThrIlySalPrometCileAspH 76
279 TTCTGTGTGTGCGCTTAACGCGGTGGCGCATTTGGCGCATCAATATA 328
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
76 eSerValIleSerArgAsnGlyValAlaIleValGlyAspGlnTYrI 93
329 TTGTGAGCTGGCATTAACGGCGGCTATACACGTTGATTTGTCGCG 378
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
93 leValSerValAlaHisAsnGlyIlyTyraSnpValAspPheGlyAla 109
379 GAAGGAGAAATCCGATGACGAGCGTTTCTTACCAAAATGTGAAAAG 428
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
110 GluGlyArgAsnProAspGlnIleSarPheThrIlyGlnIleValIly 126
429 AATATATTAAGCCCT...GACATTTACACCCCTTACAAAGCGGATTANC 475
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
126 gAsnAsnTYrGlnAlaIleArgIlyAsnIlyProTYrAspGlyAspTYr 143
476 ATATGCCGCTTTCATTAATTTGTACAGATGACAGAACCTGTCCAAATG 525
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
143 ISeMetProArGleuHisIlySnpValThrGlnAlaGlnProValGlyMet 159
526 ACAGAGCATGAGGGGAAATACCTATTCGGATTAAGAAAATATGCCGA 575
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
160 ThrThrAsnMetAspGlyIlyValIlyAlaAspArgGluAsnTYrProG 176
576 GCGTGTCCCATCGGCTCAGACACCACTATTGGCGTTATGATGATGACA 625
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
176 uArgValArgIleGlySerGlyArgGlnTYrTrpArgThrAspIlyAsp 193
626 AA.....CACGGCGATTTACTCTACTCGGCGCA.....TGG 657
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
193 IuGlnThrAsnValHisSerSerTYrTYrValSerGlyAlaIlyArgTYr 209
658 TTAAATGGGCGCAATACATATGACATATGCGGTTGGGAAATATAGCGGTAT 707
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
210 LeuThrAlaGlyAsnThrHisThrGlnSerGlyAsnGlyAsnGlyThya 226
708 TACTTTAGCGGCGATGTGCGCATGCCAAGCATATGCGCCCTATGCCGA 757
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
226 IAsnLeuSerGlyAsnValIleSerProAsnHisTYrGlyProLeuProT 243
758 TTGCAGGTGCGGCGAGCGACAGCGGTTCCCAATGTTTATTATGACAAA 807

```

```

243 hrgIlySerLysGlyAspSerGlySerProMetPheIleTYrAspAla 259
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
808 ACAAACTAAATGAGCTGCTCAACGAGGATTTTCAAAACCGGCTACCTTA 857
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
260 LysIlySpgGlnTrpLeuIleAsnAlaValLeuGlnThrClyHisProH 276
858 TTCGCGAGGAGAAACGTTTCCAGCTGATACCAAGATGATGTTCTACG 907
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
276 ePheGlyArgGlyAsnGlyPheGlnLeuIleArgGlnGlnTrpPheTYr 293
908 ATGACATTTACAGAGCGGATACATACCTC.....TTTGTGAACG 951
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
293 snGlyValLeuAlaValAspThrProSerValPheGlnArgTYrIlePro 309
952 CGCAGTACGAGCATTTTCTTACATCCACACACAGCGTACGGT... 999
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
310 ProIleAsnGlnHisTYrSerPheValSerAsnAsnAspGlyThrGly 326
1000 ...ACGTTAACAGAAACCAAGAAAGTTTCCAAATCCAAAGCTTAAG 1045
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
326 sleuThrLeuThrArgProSerIlyAspGlySerIlyAlaIlySserGln 343
1046 TACAGACAGTCCGACTGTTGACGAATCTTGAATGAATGAATGAATGA 1095
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
343 alGlyThrValLysLeuPheAsnProSerLeuAsnGlnThrAlaIlyGln 359
1096 CCAGTTTACGCGGCGGCGGTATTAATCACTACCTGCTTCAAGTTAAACA 1145
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
360 HisValIlySalAlaIleAlaGlyTYrAsnIleTYrGlnProArgMetGly 376
1146 CGGTGAACACCTTCTTTATCGATTACGGCAACGCGCAACTATCTTAT 1195
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
376 rGlyLysAsnIleTYrLeuGlyAspGlnGlyIlyGlyThrLeuThrIle 1245
1196 CAACACATCAACCAAGCGCGGCGGTTGATTTGAAGTGATTTT 1245
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
393 IuAsnAsnIleAsnGlnGlyAlaGlyIlyLeuTYrPheGlnGlyAsnPhe 409
1246 ACGGTC...TCGCTGAAACCAACGAAACGTGGCAAGCGCGGCGCTTCA 1292
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
410 ValValIlySglIlySglnAsnAsnIleThrTrpGlnIlyAlaGlyVal 426
1293 TATCATGTAAGACAGTACCTTACTTGGAAAGTAAACGCGGTGCAACG 1342
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
426 rIleGlyGlnAspAlaThrValGlnTrpLysValHisAsnProGluAsn 443
1343 ACGGCTGTCCAAATGCGCAAGGCAAGCGGTCGACGTTCAAGCCAAAGG 1392
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
443 spArgLeuSerLysIleGlyIlyLeuTYrThrLeuValAsnGlyIlySgl 459
1393 GAAACCAAGGCTCGATCAGCGTGGGCGGACGATCAGCTCATTTTGATCA 1442
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
460 LysAsnLeuGlySerLeuSerAlaGlyAsnGlyIlyValIleLeuAspG 476
1443 GCAGGACAGCATTAAGGCAAAACAAAGCCTTATAGTGAATGGCTGTCN 1492
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
476 nGlnAlaAspGlnIleGlyIlyGlnIlySglnAlaPheLysGlnValIly 493
1493 TCAGCGGAGGAGGTACGATGCACTGAATGCGCATATCAAGTCAACCCC 1542
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
493 alSerGlyArgAlaThrValGlnLeuAsnSerThrAspGlnValAspPro 509
1543 GACAAACTCTATTTGCGCTTTCGGGCGGACGTTTGATTTAAACGGGCA 1592
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
510 AsnAsnIleTYrPheGlyPheArgGlyIlyArgLeuAspLeuAsnGly 526
1593 TTCGCTTTCGTTCCACCGTATTCAAATACCGATGAAGGCGGCTGATG 1642
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
526 sSerLeuThrPheLysArgIleGlnAsnThrAspGlnGlyAlaMetIle 543
1643 NCNATCATATGTCACAAACATCCACCGTTTACATTACAGGATGAA 1692
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```


[illegible]

832	lleugllylalaasneupheglythrlllelnseratgelyasnserg	849
2456	catttaattctaaagcaaacgccccgcaaaaacggcagctgcgctttcc	2505
849	lnvalaargleuthrnguasn.....	855
2506	gacaaacgctgaagcacaacgaatgacattccgacactcaacggcaatgtctc	2555
855	855
2556	cctagccgaatgaagcagatattccattttgaaaaacacccgtttaccggac	2605
855	855
2606	actacacggcgacacagannacagactttacacttaaaagacggcaatgc	2655
856	858
2656	acgctgcgctgaagcagcaggaatttagcaattttaacstttgacacggcac	2705
859	hlsleuthrcllysnserasnryalahtsglnleuaprlleaalaargvalasp	875
2706	cattacactgaattccgcttatccgacagatgctgcacggcgcaaacgg	2755
875	sllehlslleuasnseralalaaspasnsatn.....	885
2756	gcagngtgcacacacgccccgccccgttccgcttatattatcc	2805
886	886
2806	gtttacacccgcaacttccgttagaattccggtttcaacacgctgacggataa	2855
887	valthr.....	896
2856	cgcaaatggaacngtcagaaacaaatccggttatgtgcgaactcttgc	2905
896	n...serleusergelyasnclserphenylytyrleuthrslaspleusera	912
2906	gcttacggaagcacaattgaagctggcggaagaattccgaagaaacttac	2955
912	snlysglnsllyaspysvalvalthrlylseralathrclyasnphe	928
2956	acctggcggtcaacaattacggcgacagacccgttaagccttcgaatt	3005
929	thrleuglnvalalaspysrthrcllyslupro...asnhsasnglule	944
3006	gaacgtagtaggaaggaagaaacacaaacacccgtgcgcaaaacttaatt	3055
944	utrrleupheasprala.....serlyalaaglnarqasphlsleuasn	959
3056	tcacctgcgaacacgaacacgtcgatcgccggcggtggcggtttaccaactc	3105
959	alserleuvalalysnthrvalaspleuglialatrlplystlyslleu	975
3106	atccgcaaaagacggagagtgccgctgataatccggtcaaaagaca..	3153
976	atqasnsvalasncllyarqtyraspleuylrasnprogluvalglulysar	992
3153	3153
992	gaasnglnthrvalasprthrtrasnllerthrproasnslleglna	1009
3154	3177
1009	laspyalproservalproserasnangluilllealaargvalasp	1025
3177	3177
1026	glualaprovalproproproalaproalathrproserghuthrtrgl	1042
3178	3211

1042	utpInValaIagIlaAsnSerLySInGInuBerLySthrValaIgluYsAnG	1053
3212	CGCAAGCCTTGACGGCGTGATTCGGCGCGCGCGATCCGGCGAA...	3258
1059	IuGInSpaIaThrGluThrThrAlaGInAsnAArgIuValAlaLySglu	1075
3259AAGACAGAAAGCGTGGCCGACG	3281
1076	AlaLySerAsnValLySAlaSnThrGInThrAsnGluValAlaGIn..	1091
3282	GGCCCGGCGACGCGGGGAAAAATCGCATTAATGCAGCGGAG...	3327
1092SerGlySerGluThrLySgluThrGInThrGluThrL	1105
3328GAAGAGAAAAAGCGGTGCAGCGGAT	3354
1105	ysgluThraIaThrValaIgluYsGInGluLySAlaLySValaIgluThrGlu	1121
3355	AAAGACAGCGCNTTGGCGAAA.....CAGCCGGAAGCGGAACCG	3395
1122	LySThrGInGluValaIProLySValaThrSerGInValaIserProLySgluGInI	1138
3396	GCCGNTACCACCGCTTCCCGCGCGCGCGCGCGCGGATTTGC	3445
1138	uGInSerGluThrValaIgluProGInAlaGluProAlaArgGluAsnAsp	1155
3446	CG.....CAAGCGAGCCCGAA.....	3452
1155	roThrValaSnIleLySgluProGInSerGInThrAsnThrThAlaAsp	1171
3463CCGCAACST.....	3471
1172	ThrGluGInProAlaLySgluThrSerSerAsnValaIgluGInProValTh	1188
3471	3471
1188	rgIuSerThrThrValaSnThrGlyAsnSerValaIgluAsnProGluA	1205
3472CAAGCC.....	3477
1205	snThrThrProAlaThrThrGInProThrValaSnSerGluSerSerAsn	1221
3478	CAAGCGCAAGCGC.....	3489
1222	LySProLySAsnAArgIAsArgSerValaArgSerValaProIAsnAsnVala	1238
3490G	3490
1238	IgluProAlaThrThrSerSerAsnAspArgSerThrValaIAlaLeuGlyS	1255
3491	ACCTGATNAGCGGTTATGCCAATAGCGGTTTGAGTGAATTTCCGCC...	3537
1255	spIeuThrSerThrAsnThrAsnAlaValaIleuSerAspAlaArgAlaLyS	1271
3538ACGCTCAACAGCGTTTCCCGCTACAGACGAAATTTGA	3575
1272	AlaGInPheValaIAlaLeuAsnAlaIcLySAlaValaISerGInHisIleSe	1288
3576	CGCGGTG...TTTGCGAAGACCGCGCGCAAGCGAGTTTGACAGACGCA	3622
1288	rgIuLeuGInuMetAsnAsnGluGInLyGInTyAsnValThrPaIserSof	1305
3623	TCCGGNACCAACAAACATACCGCTTGGCAAGATTTCCGCGCTACGGCCAA	3672
1305	hSerMetAsnLySAsnTySerSerSerGInTyArgAlaGluPheSerSer	1321
3673	CAAAACGACCTTGCGCAATAGTATGAGAGAAAACCTCGCAGCGG...	3720
1322	LySerThrGInThrGInLeuGInTyThrPaIserIleHisSerAsnAsnVala	1338
3721	CGCGTGGCATCCGTGTTTGCAACAACCGGACCGAAACANTCTGCAGC	3769
1338	IgluIleuIcLyGluValaIleuThrTyValaArgAsnSerSerAsnAsnPaIser	1355

```

3770 ACGGCGATCGGCACTCGGACGGCTTGGCCACGGCGCGTTTGGGGCAA 3819
      ::::::::::::::::::::
1355 ysAlaThrSerLysAsnThrLeuAlaInValAsnPhetYSerLysTyr 1371
      ::::::::::::::::::::
3820 TACGGCATCGGCGAGTTCGACATCGGCATCGACACGCGCGGGGCTTTAG 3869
      ::::::::::::::::::::
1372 TyrAlaAspAsnHisTyrTrpLysGlyTLeuSLeuGlyTyrGlyLysPhe 1388
      ::::::::::::::::::::
3870 CAGCGGCANTCTNTCAGACGATCGAGGAGGAAATCCGCGCGCGTGC 3919
      ::::::::::::::::::::
1388 eGlnSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThr 1405
      ::::::::::::::::::::
3920 TGCATTACGGCATTCAGCGACGATACCGCGCGCGTTTGGCGGATCGGC 3969
      ::::::::::::::::::::
1405 LglInPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPhetGly 1421
      ::::::::::::::::::::
3970 ATCGAAACGCTACATCGGCGCAACGGCGATTTGTCGCAAAAGCGGATTA 4019
      ::::::::::::::::::::
1422 LLeuThrProLysValGlyValArgTyrSerLysTyrLeuSerAsnAlaAsp 1438
      ::::::::::::::::::::
4020 CCGCTACGAAACGCTCAATATCGCCACCGCGCTTGCCTTCACCGCT 4069
      ::::::::::::::::::::
1438 eAlaLeuAspGlnAlaArgLysValAsnProLysSerValLysThr 1455
      ::::::::::::::::::::
4070 ACCGAGCGGGCATTAAGCGCATTTATTCATTCAAACCGCGCACACATN 4119
      ::::::::::::::::::::
1455 LArPheAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyLysPhe 1470
      ::::::::::::::::::::
4120 TCCATCACACGCTTATTNNAGCCTGCTCATACCGGATGCCGCTGCGGCAA 4169
      ::::::::::::::::::::
1471 SerValThrProLysLeuSerAlaArgTyr...AspAlaAsnGlnGlySe 1486
      ::::::::::::::::::::
4170 AGTCGGAACCGCGTCAATATCCGCGATTTGGCTCAGGATTTTCGCAAAA 4219
      ::::::::::::::::::::
1486 rGlyLysTLeuAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnG 1503
      ::::::::::::::::::::
4220 CCCGCGATCGCGAATGGCGCGTAACCGCGAATCAAGGTTTCACGCTG 4269
      ::::::::::::::::::::
1503 LngLglnInTyrAsnAlaGlyLeuLysTyrHisAsnValLysLeu 1519
      ::::::::::::::::::::
4270 TCCNTCCACGATCGCGCGCGCAAGCGCAATCGAATGGAAGCGCACACAG 4319
      ::::::::::::::::::::
1520 SerLeuLleGlyGlyLeuThrLysAlaLysGlnAlaGlyLysGlnLysTh 1536
      ::::::::::::::::::::
4320 CGCGGCGCATCAATTAAGGCTAC 4341
      ::::::::::::::::::::
1536 rAlaGlnLeuLysLeuSerPhe 1543
      ::::::::::::::::::::
seq_name: pir2:S61314

```

```

seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HF13
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61314
R:Limholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria
A:Reference number: S61314; MUID:95302961
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: EMBL:X82474; NID:g732873; PIDN:CA57857.1; PID:g732874
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
Quality: 1551.50      Length: 1754
Ratio: 1.730          Gaps: 51

```

```

Percent Similarity: 51.140      Percent Identity: 26.910
alignment_block:
US-09-303-518D-651 x S61314
Align seg 1/1 to: S61314 from: 1 to: 1561

```

```

49 AAACCGCGCGCATTCGGCTTCGCGCTGCTTACTTACGATATGCGTGC 98
      ||||| ||||| :::::::::::::::::::: ||||| |||||
2 LysThrLysArgPheLysLeuAsnAlaLysSerLeuPheLeuL 18
      ::::::::::::::::::::
99 GTTCGGCATTCCTCCCAAGCTTGGCGGACACACTTATTTTCGGCATCA 148
      ::::::::::::::::::::
18 eTyrAlaLeuThrProTyrSerGlnAlaLeuValArgAspArgVal 35
      ::::::::::::::::::::
149 ACTACCAATACATTCGCGACTTTCGCGAAATTAAGCGCAATTTGACATC 198
      ::::::::::::::::::::
35 sPyrGlnLlePheArgAspPheAlaGlyAsnLysGlyLysPheArgVal 51
      ::::::::::::::::::::
199 GCGGCGAAAGATATTTGAGCTTACAAACAAAAGGAGGAGTGTGCGCAA 248
      ||||| ||||| :::::::::::::::::::: ||||| ::::::::::::::::::::
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnLleLys 68
      ::::::::::::::::::::
249 ATCAATGCAAAAGCCCGCATGATGATTTTCTGTGCTGCGCTAAC 297
      ::::::::::::::::::::
68 nAlaLeuSerAsnValProMetLleAspPheSerValAlaAspValAsn 85
      ::::::::::::::::::::
298 .GGCGTGGCGCATTTGCTGGCGCATCATTTTGTGACGCTGCGCAT 345
      ::::::::::::::::::::
85 ysArgLleAlaThrValValAspProGlnTyrAlaValSerValLysHis 101
      ::::::::::::::::::::
346 .::::::::::::::::::::::::::AACGGGCGCTATACAACTGTA 368
      ::::::::::::::::::::
102 ALAlaLysAlaGlyValHisThrPheTyrTyrGlyGlnTyrAsnGlyHis 118
      ::::::::::::::::::::
369 TTTTGTGCGGGAAGGAAGNAATCCGATCAGACCGCTTTTCTTACCAA 418
      ::::::::::::::::::::
118 nAspValAlaAspLysGlnAsn.....GluTyrArgY 129
      ::::::::::::::::::::
419 TTGTGAAAGAAATATTAATTAAGCCTGACAAATTCACACCTTACAAAGCG 468
      ::::::::::::::::::::
129 aValGlnGlnAsnAsnTyrGlnProHisLysAlaTrpSerAlaSerAsn 145
      ::::::::::::::::::::
469 .::::::::::::::::::::::::::GATTANCAATATCCGCGTATGATTAATTTGTCAC 503
      ::::::::::::::::::::
146 LeuGlyArgLeuGlnAspTyrAsnMetAlaArgPheAsnLysPheValTh 162
      ::::::::::::::::::::
504 AGATGCAAGACCTGTCGAAATGACGAGTGCATGAGGGGG...AATACCT 550
      ::::::::::::::::::::
162 rGluValAlaProLleAlaProThrAspAlaGlyGlyLeuAspThrT 179
      ::::::::::::::::::::
551 ATTCCGATTAAGAAAATATCCGAGCGTGTCCGCGATCGGCTCAGGACAC 600
      ::::::::::::::::::::
179 yLysAspLysAsnArgPheSerPheValArgValGlyAlaGlyArg 195
      ::::::::::::::::::::
601 CACTAT.....TGGCGTTATGATGATGACAAACACG 632
      ::::::::::::::::::::
196 GlnLeuValTyrGlyLysGlyAlaTyrHisGlnGlnLysGlnLysGln 212
      ::::::::::::::::::::
633 C.....GATTATCCTTACTCCGCGGCGATGTTAATTTGGCGGC. 669
      ||||| ||||| :::::::::::::::::::: ||||| |||||
212 yTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaLleAlaGlyT 229
      ::::::::::::::::::::
670 .::::::::::::::::::::::::::AATACACAT..ATG 681
      ::::::::::::::::::::
229 hrProTyrLysAspLleAsnLleAspGlnThrMetAsnThrGlnGlyLeu 245
      ::::::::::::::::::::
682 CAGGTTGGGGAATATATGCGGTANTAGTTTGAGCGCGCGATGCGGCA 731
      ||||| ||||| :::::::::::::::::::: ||||| |||||
246 LLeuGlyPheGlyAsnHisAsnThrHisTyrSerAlaGlnGlyLeuLysGln 262
      ::::::::::::::::::::
732 TGCCACGACTATGCGCCTATGCGGATTCGACAGTGGCGGACGAGCG 781
      ||||| :::::::::::::::::::: ||||| |||||

```


1295 hrGlyAlaSerLeuThrArgHisIleAlaGlnLysSerArgAlaAspAla 1311
 3601 AACGGNTTGGACAAAGCGCATCGGACCAACCAACCTACCG 3644
 1312 GlnLysAsnSerValIlePheSerAsnIleLysTrpGlyArgAspTrpAl 1338
 3645 TTCCAGATTTCCGCGCTACCGCCCAACCAACCGACCTGCGCAATTCG 3694
 1328 aserAlaGlnIleThrArgPheSerSerLysArgThrGlnThrGlnIleG 1345
 3695 GTATGCAAAAACCTGCGCAGCGG...CGCGTCGCATCTGTTTCG 3741
 1345 LysAspArgSerLeuSerGlnLysMetGlnIleGlyValLeuThr 1361
 3742 CACAACCGGACGAAACANCTTCAGACGCGCATTCGGCACTCGGACG 3791
 1362 TySerAspSerGlnHisThrPheAspGlnAlaSerGlyLysAsnThr.. 1377
 3792 GCTTGCCACGCGCGCGTTTTCGGCAATACGCGCATCGCG...AGTTTCG 3838
 1378 .PheValGlnIleAsnLeuTrpGlyLysTrpLysLeuAsnAspIleTrp 1394
 3839 ACATCGGATACGACGCGCGCGTTTTCAGACGCGCATCTTCAGAC 3888
 1394 TrpValAlaGlyAspIleGlyLysLeuArgSerArgLeuGlnThr 1410
 3889 GGCAATCGGACGAAATCCGCGCGCGTGCATTCAGGCAATTCAGGCG 3938
 1411 GlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuThr 1427
 3939 ACGATACCGCGCGGTTTCGGGATTCGCGCATCGAACCGTCATCGCG 3988
 1427 wGlyAsnThrLeuLysIleAsnGlnPheGlnIleValProSerAlaGly 1444
 3989 CAACGCGTATTCGTCCAAAAACGCGATTCGCGTACGAAACGTCAT 4038
 1444 LeuArgTrpSerArgLysSerSerAlaAspTrpLysLeuGlyAsnAsp 1460
 4039 ATCCGACCCCGGCTGCTTCACCGGNTACCGGCGGCGCATTAAGCG 4088
 1461 ValLysValSerSerMetSerValLysThrLeuThrAlaGlyLeuAsp 1477
 4089 AGATTTATTCATCAACCGCGCAACATTCATCAACCCATTATTNA 4138
 1477 eAlaTrpArgPheLys...ValGlyAsnLeuThrValLysProLeu 1493
 4139 GCCGTCTCTAT...ACGATGCCGCTTCGGCAAGTCCGACACGCGTC 4185
 1493 eTrAlaIleTrpPheAlaAsnTrpGlyLysGlyValAsnValGly 1509
 4186 AATACCGGNTGATGGCTCAGATTTTCGCAAAACCGGACAGTGGGAATG 4235
 1510 AsnSerPheAlaTrpLysAlaAsnGlnGlnIleTrpSerAla.... 1524
 4236 GGGCGTAAACCGCAAAATCAAGGTTTCACGCTGCTCCATCCGCGCG 4285
 1525 .GlyAlaIleLeuLeuTrpArgAsnValThrLeuAsnValAsnLys 1541
 4286 CCGCAAAAGGCGCACTGGAACGCAACACGCGCGCATCAAAATTA 4335
 1541 LeuTrpLysGlyLysGlnLeuGlnLysSerGlyGlnIleLysIle 1557
 4336 GGCTACCGCTGG 4347
 1558 GlnIleArgPhe 1561

seq_name: p1r2:A37023

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
 N/A: Alternate names: immunoglobulin A1 protease type 1
 C: Species: Haemophilus influenzae
 C: Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000

C:Accession: A37023
 R:Poulsen, K.; Boudg, J.; Hjorth, J.P.; Thøgersen, H.C.; Kilian, M.
 Infect. Immun. 57, 3097-3105, 1989
 A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophilus influenzae
 A:Reference number: A37023; MUID:89379374
 A:Accession: A37023
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1541 (POU)
 A:Cross-references: GB:X64357; NID:q43560; PIDD:CAA45708.1; PID:q43561
 A:Experimental source: serotype b
 C:Superfamily: IgA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1536.50 Length: 1722
 Ratio: 1.677 Gaps: 56
 Percent Similarity: 53.194 Percent Identity: 26.887

alignment_block:

US-09-303-518D-651 x A37023 ..

Align seg 1/1 to: A37023 from: 1 to: 1541

64 CGCTTCGCGCTCTTACGATATGCGCTGCGCGCATTCCTTC 113
 5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTrpAlaLeuThr 21
 114 CCAAGCTTGGCGGACACATTTATTCGGCATCAATCAATCAATATC 163
 21 oGTrpGlnAlaIleAlaLeuValArgAspAspValAspTrpGlnIlePhe 38
 164 GCGACTTGGCGGAAATTAAGCAATTCAGTTCGCGCGCGCAAGATAT 213
 38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
 214 GAGGTTCACCAAAAGGAGGAGTGTGCGCAATTCAGCAAAAGC 263
 55 LeuValLysAspLysAsnLysAsnLysAspLeuGlyThrAlaLeuProAsn 71
 264 C...CCGATGATTTGATTTCTGCGTGC...CGTAAACGCGTGGCG 307
 71 yIleProMetIleAspPheSerValAlaAspValAspLysArgIleAla 88
 308 CATTCGCGGCGCATCAATATTCGACGCGTGCATAT...AAGCGCGC 354
 88 hrLeuIleAsnProGlnTrpValValGlyValLysHisValSerAsnGly 104
 355 TATTAACAGCTGATTTGGT.....GCGGAAGGAAGNA 389
 105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnLys 121
 390 TCCGATCAAGCAGCT.....TTTCTTCACAATTC 421
 121 palLysAlaIleArgAspValSerSerGlnIleAsnArgTrpPheSer 138
 422 TGAAGAATAATATAT.....AAG 441
 138 alGlnLysAsnGlnTrpProThrLysLeuAsnGlyLysThrValThrThr 154
 442 CCTGCAATTCACACCCCTTACACGCGGATTAACGCGGCTTTCGCA 491
 155 GlnAspGlnThrGlnLysArgValGlyAspTrpTrpLysMetProArgLeu 171
 492 TAAATTTGTCAGATGACGAACCTGCGAA...ATGACAGTGCATGA 538
 171 rLysPheValThrGlnValAlaProIleGlnIleAlaSerThrAlaSerSer 188
 539 GGGGCAATCTTTCGATTAAGAAATATTCGCGAGGCTGCGCATC 588
 188 spAlaGlyThrTrpAsnAspGlnAsnLysTrpProAlaPheValArgLeu 204

[illegible]

757 G1yArGsnaValAlaAsn1LeThrSerAsn1LeThrAlaSerAsn1LysAl 773
2212 CCAAGCGATTTTGGTGTGGCACCAGCATCAAGCCATACATCTGTGTAC 2161
773 aGlnValHis1LeG1yTyr.....LysThrG1yAspThrValAlcysValA 788
2162 GTTTCGACGTGACNGGCTGTGCACAAATTGTGTGCANNAANCAATTTCCGAC 2211
788 rGserAspTyrThrG1yTyrValThrCysThrThrAspLysLeuSer... 803
2212 GATTAAGTGTCTCTCATTTGACTTAAGACNGACNCTNAGCGGCATGTNAG 2261
804 AspLysAlaLeuAsnSerPheAsnProThr..... 813
2262 NCTNNCNA1TNAAGTNNNTTNAANCTCNCGGCGTTCNNACNMANAG 2311
813 813
2312 GCATCTTGTAGTGCAAATGGCGATACAGTTATACAGTACGACCAAGGCC 2361
814 ..AsnLeu..... 815
2362 ACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAA 2411
816ArgG1yAsnValAsnLeuThrG1uSerAlaAsnPheValGlu 830
2412 TCAGACCAATTAACGCGACACNCAATCGGNTTGGGCAATGCTTCATTTA 2461
830 YLysAlaAsnLeuPheG1yThr1LeG1nSerArgG1yAsnSerGlnValA 847
2462 ATCTAAGCAACAAAGCGCGCAACAAAGCGAGCTGACGCTTCCGACAAAC 2511
847 rG1euThrG1uAsn..... 851
2512 GCTAAGCAACAGTAAAGCCATTCGCGACTCAACGCGAATGTCTCCTAGC 2561
851 851
2562 CGATTAAGCGCATTTTCATTTTGGAAAACAGCCGTTTACCGGACATCTCA 2611
851 851
2612 GCGGCAAGCAAGANACAGCATTAACAAAGACAGGANTGACGCGTG 2661
852SerHisThrPheLeu 856
2662 CCGTCAGGACAGGAATTAAGCAATTTAAACCTTGACAAAGCCACCATTTAC 2711
857 ThrG1yAsnSerAspValHisG1nLeuAsnPheValAsnG1yHis1LeuH 873
2712 ACTCAATTCGCGCTTATCGCACAGATGTCGAGGGCGCAAAACCGCGCAGNG 2761
873 s1euAsnSerAlaAspAsnSerAsn..... 881
2762 TGTACAGACAGCGCGCGCGCTTCGCGCGCTTCCTATTATCCGTTACA 2811
882AsnValThr 884
2812 CCGCAACTTGGTAGAATCCGTTTTCACACGCTGACGAGTAAACGGCA 2861
885LysTyrAsnThrLeuThrValAsn...Se 893
2862 ATTGAACNGTCAAGACATTCGCGTTATGTGCAACTCTTGCGGTAC 2911
893 r1euSerG1yAsnG1ySerPheTyrTyrLeuThrAspLeuSerAsnLysG 910
2912 GAAGGACAAATTGAAGTGGGCAAGTTCGGAAGNACTTACACCTTG 2961
910 InG1yAspLysValValThrLysSerAlaThrG1yAsnPheThrLeu 926
2962 GCGGTCAACAATACCGGCAACCCGTTACGCTGCATCATTTGACGG 3011
927 GlnValAlaAspLysThrG1yGluPro...AsnHisAsnG1nLeuThrLe 942
3012 AGTGAAGGGAAGACACAACACCGCTGTCCGAAACCTTAATTCACCC 3061
942 uPheAspAla.....SerLysAlaGlnArgAspHis1LeuAsnValSerL 957
3062 TGCAAAACGAACAGTGTGATGCCGGCGGCGGTACCAATCTACATCCCG 3111
957 euValG1yAsnThrValAspLeuG1yAlaThrP1yLysTyrLysLeuArgAsn 973
3112 AAAGACGCGAGTTCCTCGCATTAATCCGGTCAAAAGAACAA..... 3153
974 ValAsnG1yArgTyrAspLeuTyrAsnProGluValG1uLysArgAsnG1 990
3153 3153
990 nThrValAspThrThrAsn1LeThrThrProAsnAsn1LeGlnAlaAspV 1007
3154GAGCTTCCGACAAATCGCGCAG..... 3177
1007 a1ProSerValProSerAsnAsnG1uGln1LeAlaArgValAspGluAla 1023
3177 3177
1024 ProValProProProAlaProAlaThrProSerG1uThrThrG1uThrVa 1040
3178 .GCAGAAAGCCAAAACAGCGCGGAA.....GACAAAGCGCGCAA 3217
1040 lAlaG1uAsnSerLysG1nSerG1ySerThrValG1uLysAsnG1uGlnA 1057
3218 GCGTTGACGCGCTGATTTGCGCGCGCGGAGATCGCGCGAA..... 3258
1057 sPalaThrG1uThrThrAlaGlnAsnArgG1uValAlaLysG1uLalys 1073
3259AAGACGAAGCGTTCGCCGAACCGCGCCG 3287
1074 SerAsnValLysAlaAsnThrG1nThrAsnG1uValAlaGln..... 1087
3288 GCNGCAGGCGGGAATGTCCGATTAATGACAGCGGAG..... 3327
1088 ...SerG1ySerG1uThrLysG1uThrG1nThrThrG1uThrLysG1uT 1103
3328GAAGAGAAAAAGGGGTGCAGCGGATTAAGAC 3360
1103 hrAlaThrValG1uLysG1uGluLysAlaLysValG1uThrG1uLysThr 1119
3361 AGCGCTTGGCGAAA.....CAGCGCAAGCGGAACCGCGCGGN 3401
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluGlnSe 1136
3402 TACCACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCG... 3447
1136 rG1uThrValGlnProGlnAlaG1uProAlaArgG1uAsnAspProThrV 1153
3448CAACGCAAGCCCGCAA..... 3462
1153 alAsn1LeuLysG1uProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1169
3463CCGCAACT..... 3471
1170 GlnProAlaLysG1uThrSerSerAsnValG1uGlnProValThrG1uSe 1186
3471 3471
1186 rThrThrValAsnThrG1yAsnSerValValG1uAsnProG1uAsnThrT 1203
3472CAACCG.....CAACCG 3483
1203 hrProAlaThrThrGlnProThrValAsnSerG1uSerSerAsnLysPro 1219
3484 CAGCGC..... 3489
1220 LysAsnArgHisArgArgSerValArgSerValProHisAsnValG1uPr 1236


```

3490 ..... GACCTGA 3496
1236 calatThrThrsersAsnspargSerThrValAlaLeuCyAspLeu 1533
3497 TNACCGCTTATGCAATAGCGCTTGGATGATTTCCGCC ..... 3537
1233 hrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaValAlaGln 1269
3538 ..... ACCTCAACAGCGTTTCCCGCTACAGCAGCAAAATGGACCGCT 3581
1270 PheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1286
3582 G...TTTCCCAAGACCGCCCAACGCGNTTGGACAAACGACATCCGGN 3628
1286 uGlnMetAsnAsnGlnGlyGlnTyrAsnValTyrValSerAsnThrSerM 1303
3629 ACACCAACATACCTACCGTTCGCAAGATTTCCGCTACCGCCCAACAAAC 3678
1303 etAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSer 1319
3679 GACCTGCGCCCAATCGGTATGCAGAAAACCTCGGACGCGG...CGCGT 3725
1320 ThrGlnThrGlnLeuGlyTyrPaspGlnThrIleSerAsnAsnValGlnLe 1336
3726 CGGCATCCTGTTTTCGCAACACCGGACGCAAAACANTTCGACGACGCA 3775
1336 uGlyGlyValPheThrTyrValArgAsnSerAsnAsnPhesAspLysAla 1353
3776 TCGGCACTCGGACGCGCTTCCGCGCGCGCTTTTCGGCAATACGCG 3825
1353 hrSerLysAsnThrLeuAlaGlnValAsnPhetyrSerLysTyrTyrVala 1369
3826 ATCGGACGAGTTCACATCGGCATCAGCAGCGCGCGGCTTTTACAGCGG 3875
1370 AspAsnHisThrTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSe 1386
3876 CANTCTNTCAGACGGCATCGGAGGCAAAATCCGCGCGCGCTGCTGCATTT 3925
1386 TrpLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnP 1403
3926 ACGGCATTCAGGACGATACCGCGCGGCTTTCGGGAGTTCGGCATTCGAA 3975
1403 heGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyLysAsnPhetGlyLeThr 1419
3976 CCGTTCATCGCGGCAACGGGCTATTTCGTCGCAAAAGCGGATACCGCTA 4025
1420 ProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLe 1436
4026 CGAAACGCTCAATATGCGCCACCGCGGCTTCGCTTCAACCGGATACCGG 4075
1436 uAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaPheA 1453
4076 CGGGGATTAAGCAGATTATTCATCAACCGGCGCAACACATNTCCATC 4125
1453 IeGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSerVal 1468
4126 ACNCGTTATTNAGCCTGCTCTATACCGATGCCGCTTCGGGCAAGTCCG 4175
1469 ThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGlyLys 1484
4176 AACACGCGCTATACCGCGNGTATTTGGCTCAGGATTTCCGCAAAACCCCA 4225
1484 sTleAsnValAsnGlyTyrAspPheAlaLysTyrAsnValGluAsnGlnGln 1501
4226 GTGCGGAATGGGCGTAAACGCGCAATCAAGGTTTTCACGCTGTCCTTC 4275
1501 IntTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1517
4276 CACGCTGCCGCGCCAAAGNCCGCAACTGAGACGCAACACAGCGCGG 4325
1518 IleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaGln 1534
4326 CATCAATTAAGCTAC 4341

```

```

1534 uLeuLysLeuSerPhe 1539
seq_name: p1r2:A81018
seq documentation block:
serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogr
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81018
R:Reteljin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rt, H.; Qin, H.; Yamathewan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: A81018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <TEXT>
A:Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAF42325.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1998
alignment_scores:
Quality: 1522.00 Length: 1639
Ratio: 1.770 Gaps: 51
Percent Similarity: 52.471 Percent Identity: 27.639
alignment_block:
US-09-303-518D-651 x A81018 ..
Align seg 1/1 to: A81018 from: 1 to: 1431
55 GGGCGGATCGGCTTTCGCGCTGCTTACTTACCATATGCTGCTGCG 104
17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValaArgAsn..... 31
105 CATTCCTCCCAAGCTTGGCGGACACACTTTCGGCATCAACATACC 154
32 .....AspValAspTyrG 36
155 AATATCTCGGACTTTCGCGAAATTAAGCAAGTTTGCAGTCGGGCG 204
36 IntTyrPheArgAspPheAlaGlnAsnLysGlyAlaPheThrValAla 52
205 AAGATATTGAGGTNTTCAACAAAGGAGGAGTTCGTCGCAATCAAT 254
53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69
255 GACAAAGCCCGCATGATGATTTTCTGCGTGCCTGCT...AACGGCG 301
69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAla 86
302 TGGCGGATTTGGTGGGATCAATATATTTGTAGCGTGCACATAACGCC 351
86 IeAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102
352 GGCATATAACAAGTTGATTTTGGTGGGAGGAAGNAAATCCGATCGCA 401
103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGlnGln 119
402 CCGTTCCTTACCAATTTGCAAAAGAAATTAATTAATTAACCTGACAA 451
119 AlaTyrThrTyrArgLeuValSerArg.....Asp 130
452 CACACCTTACACGCGGATTAATCAATATGCGCGTTTGCATAATTTGTC 501
130 roHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeuVal 145

```


708LysArgGluProValLeuGluAsnG1 716
2001 CTGGATCNACCGACGCTTTAAAGCGAAATTTCCATATTTCAGGCG...G 2047
716 uTrpThrArgLysSerPheLysAlaLalArgPheThrLeuArgAsnHisA 733
2048 GCGAGCGCGGTGATTTCCGCAATGTTCSSAAAGTGAAGCGGCAATGNCAT 2097
733 LArGLeuThrAlaG1ArgAsnThrAlaHisLeuAspG1LysAspLeuThr 749
2098 TTGAGCAATCAGCGCCAGACAGTTTGTGTGCGACCCGCAATCAAGGCA 2147
750 AlaTyAspLeuSerG1LLeAspLeuG1LrheThrG1LysThr 766
2148 TACAACTGTACACGCTTCGAGCTGACNGTGTGCAAAATGTGTGCAAN 2197
766 oGlu...CysTyArgSerTyHisSerG1LysThrHisCys..... 779
2198 AAACATTTACCGACGATAAGTGATGTTCATGTGACTAAGACNACNTN 2247
779 779
2248 AGCGGCAATGTNAGNCTNNCNAATNAGNTNNTTAAANCTCNGGCGN 2297
779 779
2298 TGCNACCTNAAANGCAATCTTAGTGCAAATGCGATACACGTATACAG 2347
779 779
2348 TCAGCCACAACGCCACCCAAAAGCGCAACCTTACCTCGGCGCAATGCC 2397
780ThrProAsnAlaValLeuLys..... 786
2398 CAAGCAACATTTAATGACGCCACATTTAAAGCGCAACNATCGGNTTCGGG 2447
786 786
2448 CAATGCTTCATTTAATCTAAGACACACGCCGACAAAAGCGAGTCTGA 2497
786 786
2498 CGCTTCCGACAAACGCTAAGGCAACGTAAGCATTCGCGACTCAACGCG 2547
787AlaG1LysnTyArgAla...LeuProAlaThrG1LValAlaArgLys 800
2548 AATGTGCTCCGTAGCGGATAGGACAGATATTCATTTGAAACACAGCGCGT 2597
801 AspLeuThrLeuAsnAspArgSerG1LLeuArgLeuG1LysAlaHisLe 817
2598 TACCGGCAACTCAGCGGACGACAGGAAAGNACAGCATTTACACTTAAGA...G 2644
817 uTrpG1LysSerLLeArgAlaG1LysAspThrAlaValAlaArgMetG1LAla 834
2645 ACAGGCAATGAGAGCTGCGGTGACGACGCGAAATTAAGCAATTTAAACCT 2694
834 spSerAsnTrpThrLeuSerG1LnsSerHisThrG1LAlaLeuThrLeu 850
2695 GACAAAGCGCAATTAACACTCACTATTCGCGTATTCGCAAGATGCTGCAG 2744
851 AspG1LAlaG1LnlThrLeuAsnProAspPheAlaAsnAsn..... 864
2745 CCGCGAAACCGGACAGNGTGTGACACAGCGCGCGCGTTCGCGCGGT 2794
864 864
2795 CCTATTATCGCTTACACCGCAACTCGGTAGAAATCCGTTTCAACAG 2844
865ThrHisAsnAsnArgPheAsnThr 872
2845 CTGACGGTAAAGCGCAATGACNGTCAAGGACATTCGCTTATATGTC 2894
873 LeuThrValAsnG1LThrLeuAspG1LrheG1LThrPheArgPheLeuTh 889
2895 GGAACCTCTTC...GCTACCGAAGCGAAATTTGAACCTGCGGAAAGTT 2941
889 rG1LlLeValArgLysGlnAsnAlaProProLeuLysLeuGlnG1LAspS 906
2942 CCGAAGNACTTACACCTTGCGGCTCAACAAATACCGGCAACACCCGTA 2991
906 eArgG1LAlaPheG1LlLeHisValLysAsnTrpG1LgGlnG1LProGln 922
2992 AGCCTGCATCAATGTAGGTAGGAGGAAAGACAAACACCCGCTGTC 3041
923 ThrThrG1LSerLeuAlaLeuVal.....SerLeuAsnProLysHisSe 937
3042 CGAAACCTTAATTTCAACCTGCAAAAAGCAACGCTGATGCGCGCGCT 3091
937 rHisG1LAlaArgPheThrLeuGlnAsnG1LTYrAlaAspLeuG1LAla 954
3092 GCGGTACCAACTCATCGCGCAAGACGCGGATTCGCGCTGCATTAATCG 3141
954 YrArgTyLlLeuArgLysAsnAsnAsnG1LYrSerLeuTYrAsnPro 970
3142 GTCAAGACACAGAGCTTTCGACAACTCGCGAAGCGCAAGCAACAAA 3191
971 LeuYsG1LAlaG1LLeu.....GlnLlLeG1LAlaThrAr 982
3192 ACAGCGGAAAAAGACACACGCGCAAAACCTTGACGCGCTGATTGCGCGC 3241
982 gAlaG1LHisG1LrArgAsnGlnGlnAlaTYrAsnGlnLeuGlnAlaThr. 998
3242 GCGCGATCCCGCGGAAAGACAGAAAGCTTGCGGACCGCGCGCGCG 3291
999AspLlSerArgG1LValG1LnlAspSerAlaTrpArgGln 1013
3292 GCAGCGCGGAAAAATGTCGCAATTAATGACAGCG.....GAGCA 3329
1014 Ala.....LeuGlnAlaTrpLlAsnSerGlnTh 1023
3330 AGAGAAAAAAGCGGTG.....CAGCGGATTAAGACAGCGGCTTGCGCA 3373
1023 rG1LLeuAlaArgLlAspSerGlnValG1LYrLeuSerAlaGlnLeuL 1040
3374 AACAG..... 3378
1040 YsGlnThrAspProLeuThrG1LYrLeuThrArgAlaGlnAsnLeuYs 1056
3379CGCGAGC 3386
1057 AlaAlaGlnG1LYrSerAlaAspLlLecYsArgGlnValAlaLysAlaL 1073
3387 GGAACCCGCGCGGNTAC.....ACGCGCTTCC 3415
1073 aAspThrAsnAspLeuThrLeuPheG1LrG1LLeuAspThrTYrLlLeG 1090
3416 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGGACCGGCAACCG 3465
1090 LndArgValG1LlLeuValG1LnsSerG1LLeuAspYsAlaArgGlnG1LY 1106
3466 CAACCTCAACC..... 3477
1107 AspAlaGlnAlaValAlaG1LrAlaArgHisAlaTYrLeuAsnAlaLeuAs 1123
3477 3477
1123 nArgLeuSerArgGlnLlLeHisSerLeuYsThrG1LYrAlaAlaG1LYlLea 1140
3478 ..CAACGCGAG...CGGACCTGATNAGCGGTTATGCAATACGCGGTTTG 3522
1140 rMetProAsnLeuAlaG1LLeuLlLeSerArgSerAlaAsnThrAlaVal 1156
3523 AGTGAATTTTCCGCGACGCTCAACAGCGTTTTCGCGCTACAGACGAAT 3572
1157 SerGlnGlnAlaAlaTYrAsnThrG1LYrGlnGlnAlaG1LYrArgG1l 1173


```
121 nAlaLysSerHisArgSpValSerSerGluGluAsnArgTyrPheSerV 138
422 TGAAGAATAATATAT.....AAG 441
138 aIGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
442 CCTGACAATTCACACCTTACACGGCGATTACATATGCCGCTTTGCA 491
155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171
492 TAAATTTGCAGAGATGCAGAACCTGTGCA...ATGACGAGTGCAGATGA 538
171 pLysPheValThrGluValAlaProIleGluLysSerThrAlaSerSera 188
539 GGGGGAATACCTATTCGATAAAGAAAATATCCGAGCGTGTCCGATC 588
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
589 GGCTCAGCACACCACTATTGGCGTTATGATGACAAACAGCGGATTT 638
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp... 216
639 ATCCTATCCGGCGCATGGTAAATGGCGGCAATACACATATGACGGGT 688
217 .AsnTyrSer.....LeuIleLeuAsnAsnHisGluValGly. 228
689 GGGGAATATAT.....GGCGTANTT 708
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
709 AGTTTGAAGCGCGCATGTGCCCATGCCAACGAC.....TANCG 746
245 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheG 261
747 C.....C 748
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278
749 CTATGCCGATTCGAGGTGCGGAGCGGACGAGCGGTTCGCAATGTTTAT 798
278 roleuThrAsnTyrAlaValAlaLeuGlyAspSerGlySerProLeuPheVal 294
799 TATGACAAAACAACATTAATGCTGCTCAACGCGATTTTACAAACCG 848
295 TyrAspArgGluLysGlyLysTrpLeuPheLeuGly.....Se 307
849 CTACCTTATTCGCGAGGAAAACGGTTCCAGCTGATGACAAAGATT 898
307 rTyrAspPheTrpAlaGlyTyrAsn.....LysLysSer 319
899 GGTTCACGATGACATTTACAGAGCGGATACACATACCGTCTNTTTGAA 948
319 rPheGlnLysTrpAsnIleTyrLysProGlu..... 328
949 CCGCGCAGTACGCGACATTTTCTTACATCACACACAGCGTACG... 996
329 .....PheAlaLysThrValIleuAspLysAspThrAl 339
997 .GGTAGGTAACGAAACACGAAAGGTTC...AATCCAAAGCTTA 1042
339 aGlySerLeuThrGlySerAsnThrGlnTyrAsnTrpAsnProThrGlyL 356
1043 AAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAAT..... 1080
356 yThrSerValIleSerAsnGlySerGluSerLeuAsnValAspLeuPhe 372
1081 .....GAAACTGATTAAGAACAGATTTCGCGGCGAGGGGTGT 1118
373 AspSerSerGlnAspThrAspSerLys..... 381
1119 TATTCAGTACCGTCCAAAGGTTAAACAGGTGAAAACCTTGCTTTATCG 1168
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391
```

```
1169 ATTACGGCAACGGCAAACTCTATTCAACAACATCAACCAAGCGCG 1218
392 .ArgGlySerGlyThrLeuThrLeuAsnAsnHisLeuAspGlnGlyAla 407
1219 GGGGCTTGTATTTTGAAGGTGATTTACGGTC...TCGGCTGAAAACA 1265
408 GLyGlyLeuPhePheGlnGlyAspTyrGluValLysGlyLysThrSerSp 424
1266 CGAACGCTGGCAACGGCGCGGCTTCATATCAGTGAAGACAGTACGTTA 1315
424 rThrThrTrpLysGlyAlaGlyAlaSerValAlaAspGlyLysThrVal 441
1316 CTTGAAAGTAAACGGCGCTGGCAACGCGGCTGTCCAAATTCGGCAAA 1365
441 hrTrpLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
1366 GGCAGCGTGCACGTTTCAAGCCAAAGGGAAACCAAGCTCAGTACGCT 1415
458 GlyThrLeuIleValGlnGlyLysGlyLysAsnLysLysSerLeuLysVa 474
1416 GGGGAGCGATACAGTCATTTTGATCAGACGACGACGATAAAGCAAAA 1465
474 IGLYAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysV 491
1466 AACAGCCTTTAGTGAATGCGCTTGTACGCGGAGGAGGTAGGTCGAA 1515
491 aLysAlaPheSerGlnValGlyIleValSerGlyArgSerThrValVal 507
1516 CTGAATGCCGATATCAGTTTCAACCCGACCAACTATTTTGGCTTTCG 1565
508 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheAr 524
1566 CGCGGAGCGTTTGATTTAAACGGGCAATCGCTTTCGTCACCGTATTC 1615
524 gGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIleA 541
1616 AAAATPACCGATGAAGGCGCATGTTGNCNATCTAATGCCACAAACACA 1665
541 rGAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThr 557
1666 TTCACCGCTTACATTCACGGGATGAAGTATTACACACCGAGT..... 1710
558 SerThrValThrIleThrGlyLysSerLeuIleThrAspProAsnThrI 574
1710 ..... 1710
574 eThrProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheA 591
1710 ..... 1710
591 rGArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThr 607
1710 ..... 1710
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624
1711 .....GTAAGAAAT..... 1719
624 nSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAsp 641
1720 .....ATCATAGACTTATTTACACCAAGAAATTTGCC 1752
641 lAlaIleLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnGly 657
1753 TACAACGGTGTGTTGGGAGAGAAGATACGCAAAACGAGCGCGCT 1802
658 PheAsnGlyTyrPheGlyGlnGlnGlu...GlyLysAsnAsnGlyAsnLe 673
1803 CAACCTGTGTACAGCCCGCGCAGAGACGCGACCCGCTGCTTTCG 1852
673 uAsnValThrPheLysGlyLysSerGlnGluAsnArgPheLeuLeuThrG 690
```

```
1853 GCGGAACAATTTAAACGGCAACATCATCGCAACAAACGCAACACTGTTT 1902
|||||
690 LYLGLYHRSNLEUASNGLYASPLEUYSVALGLULYSGLYTHRLEUPHE 706
|||||
1903 TTCAGCGGACAGACCGACCGCATCAATCAATTTAGAACGGGGTG 1952
|||||
707 LEUSERGLYRGPRTHRPROHISALAARGASPILEAGLYILESESE 723
|||||
1953 GTCAAAAATGGAAGT.....ATCCCAACAGAGAAATCCGTGGGACA 1996
|||||
723 rthrlslysbpGlnHisPheAlaGluSnAsnGluValValGluVal 740
|||||
1997 AGCACTGATCAACCGCACGTTTAAAGCGGAATTTCCATATTCAGGAC 2046
|||||
740 spAspPrLLeAsnArgAsnPhelYsAlaThrAsnILeAsnValThrSn 756
|||||
2047 GGGACGGCGGTATTC...CGCAATGTTGCCAAAGTGGAGCGGATG 2093
|||||
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnI 773
|||||
2094 NFATTTGACCAATCACGGCCACAGCTTTTGTGTGCGACCGCATCAA 2143
|||||
773 ethrAlaSerAspAsnAlaLysValAlaHisIleGlyTyr.....LysAlaG 788
|||||
2144 GCCATACATCTGTACAGCTTGACGTGACNGSTGTGACAAATGTGTC 2193
|||||
788 LysPThrValLysValAlaArgSerAspTyrThrGlyValThrCysThr 804
|||||
2194 GAANAANACATTAACGACAGATTAAGTGTGTTGATTGACTAAGACNGA 2243
|||||
805 ThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsnAlaThrAs 820
|||||
2244 CNTNAGCGGCANTGTNAGNCTNNCNATNAGNTNNTNNAANCTCNCNG 2293
|||||
820 nValSer..... 822
|||||
2294 GGCNTGCNNCACTNANAGCAATCTTAGCAATGGCATACACGTTAT 2343
|||||
822 ..... 822
|||||
2344 ACAGTCAGCCACAAACGCCACCAACCAACCTTAGCTGTGGGCA 2393
|||||
823 .....GlyAsnValAsnLeuSerGlyAs 830
|||||
2394 TGCCCAACAAACATTTAATCAAGCCACATTAACGGCAACNCATCGGNTT 2443
|||||
830 nAlaAsnPhelValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847
|||||
2444 CGGGCAATGCTTCATTTAATCTAAGCAACACGCGGCACAAACGGCAGT 2493
|||||
847 hGlyAsnSerGlnValAlaArgLeuThrGluAsn..... 857
|||||
2494 CTGACGCTTTCGACACACGCTAAGCAACGTAAGCCATTCGCACTCAA 2543
|||||
857 ..... 857
|||||
2544 CGGCATGTCTCCCTAGCCGATTAAGGCACTATTCATTTGAAAAACAGCC 2593
|||||
857 ..... 857
|||||
2594 GCTTACCGGACAACACTCAGCGGACGAAGANACAGCATTTACACTTAAA 2643
|||||
857 ..... 857
|||||
2644 GACAGCGAATGAGCGCTGCCGTGACGACGCAATTAAGCAATTTAAACT 2693
|||||
858 ...SerHisThrPHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnLe 873
|||||
2694 TGACACGGCCACCATCTACATTCGCGCTATCGCCACAGATCTCTCAG 2743
|||||
873 uAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAspAlaAsnL 888
|||||
2744 GCGGCAAAACCGGACGNGTGTACAGACGCGCGCGCGCTTCCGGCGGT 2793
|||||
```

```
888 yValThrThr..... 891
|||||
2794 TCCCTATTATCCGTTACACCGCAACTTCGGTAGAATCCGTTTCAACAC 2843
|||||
892 .....TyrAsnThr 894
|||||
2844 GCTACGCGTAACGCGCAATTTGAACNGTCAAGGAACATTCGCTTATGT 2893
|||||
894 rLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910
|||||
2894 CGGAACCTTCGCTACCGAAACGCAATTTGAAGCTGGCGGAAAGTCC 2943
|||||
910 hAspLeuSerAsnLysGlnGlyAspLysValValValThrLysSerAla 926
|||||
2944 GAAGNACTTACACCTTGCGGCTCAACAAATACCGGCAACGACCGGTAAG 2993
|||||
927 ThrGlyAsnPherThrLeuGlnValAlaAspLysThrGlyLurProThrLys 943
|||||
2994 CCTGATCAATTTGACGCTAGTGGAGGAAAGACAAACGCGCTGCG 3043
|||||
943 s...AsnGluLeuThrLeuPheAspAla.....SerAsnAlaThrArgA 957
|||||
3044 AAACCTTAATTTACCCGCAAAACGACACGCGCATGCGCGCGGTG 3093
|||||
957 snAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuGlyAlaThr 973
|||||
3094 CGTTACCACTCATCCGCAAAAGACGGCGAGTTCGCTGCAATATCCG.. 3141
|||||
974 LysTyrLysLeuThrArgAsnValAsnGlyArgTyrAspLeuTyrAsnProG 990
|||||
3141 ..... 3141
|||||
990 uValGluLysArgAsnGlnThrValAspThrThrAsnIleThrThrProA 1007
|||||
3141 ..... 3141
|||||
1007 snAsnIleGlnAlaAspValProSerValProSerAsnAsnGluIle 1023
|||||
3141 ..... 3141
|||||
1024 AlaArgValGluThrProValProProProAlaProAlaThrProSerG 1040
|||||
3141 ..... 3141
|||||
1040 uThrThrGluThrValAlaGluAsnSerLysGlnGluSerLysThrValG 1057
|||||
3141 ..... 3141
|||||
1057 LuTysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnGlyLurVal 1073
|||||
3141 ..... 3141
|||||
1074 AlaGluGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGluVa 1090
|||||
3142 .....GTCAAAG 3148
|||||
1090 lAlaGlnSerGlySerGluThrGluGluThrGlnThrGlnIleLysG 1107
|||||
3149 AA..... 3150
|||||
1107 lUrThrAlaLysValGluLysGluGluLysAlaLysValGluLysAspGlu 1123
|||||
3151 .....GAAGGCTTTCGACAACTCGCGCA 3176
|||||
1124 lIleGlnGluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLys 1140
|||||
3177 G..... 3177
|||||
1140 sProAlaProLysGluValSerThrAspThrLysValGluGluThrGlnInv 1157
|||||
3177 ..... 3177
```

```

1157 aLglnAlaGlnProGlnInhrGlnSerThrThrValAlaAlaAlaGlnAla 1173
3178 .....GCAGAGCCAAAAACGGGGGAAAAAGA 3206
1174 ThrSerProAsnSerLysProAlaGlnGlnThrGlnProSerGlnLysThr 1190
3207 CACGCGCAAGAGCCCTGACGCGGTGATGGCGCGG..... 3243
1190 rAsnAlaGlnProValThrProValValSerLysAsnGlnInhrGlnAsnT 1207
3244 .....CGCATGCCCGCCGAAAAAGACAGAA 3267
1207 hrThrAspGlnProThrGlnLysThrAlaLysValAlaGlnLysThrGln 1223
3268 AGCGTGGCCGAAACCGCGCGCGGCGAGC..... 3297
1224 LysThrGlnGlnProGlnValAlaSerGlnAlaSerProLysGlnGln 1240
3298 .....GGGGAATAATGTC 3309
1240 uGlnSerGlnThrValAlaGlnProGlnAlaValLeuGlnSerGlnLysValP 1257
3310 ..GGCATATGCAAGCGGAGAGAGAAAAAGGGTGCAGCGGATAA 3357
1257 rOThrValAsnAsnAlaGlnGlnAlaGlnLeuGlnInhrGlnInhr 1273
3358 GAC..... 3360
1274 SerAlaThrValSerThrLysGlnProAlaProGlnAsnSerLysAsnT 1290
3361 .....AGCGNTTGGCGAAACAGCGGAAACCGGAAACCGCGCGG 3400
1290 rGlySerAlaThrAlaLeuThrGlnThrAlaGlnLysSerAspLysProG 1307
3401 NNAAC..... 3405
1307 InThrGlnThrAlaAlaSerThrGlnAspAlaSerGlnHisLysAlaAsn 1323
3406 .....ACGCGCTTCCCGCG 3419
1324 ThrValAlaAspAsnSerValAlaAsnAsnSerGlnSerSerAspProLys 1340
3420 CGCGCGCGCGCGCGCGCGGATTTGCCGACCGCAG..... 3456
1340 sSerArg...ArgArgArgSerLysSerGlnProGlnGlnInhrSerAlaG 1356
3456 ..... 3456
1356 InGlnThrThrAlaAlaSerThrAspGlnThrThrAlaAlaAspAsnSer 1372
3457 CCCCAACCGCAACCTCAACCCCAACCGCAG..... 3486
1373 LysArgSerLysProAsnArgArgSerArgArgSerValArgSerGlnLysP 1389
3487 .....CGCGACCTGA 3496
1389 oThrValThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuT 1406
3497 TNAGCGGTATGCCATAGCGGTGATGAGTAATTTCCGC..... 3537
1406 hrSerThrAsnThrAsnAlaValLysSerAspAlaMetAlaLysAlaGln 1422
1423 rHeValAlaLeuAsnValGlyLysAlaValSerGlnHisLysSerGlnLe 1439
3538 .....ACGCTCAACAGCGTTTCCGCTACAGAGCAATTTGACCGCGCT 3581
1439 rHeValAlaLeuAsnValGlyLysAlaValSerGlnHisLysSerGlnLe 1439
3582 G...TTTGGCGAAGACCGCGCGCAACCGCTTTGGACACGCAATCCGCGN 3628
1439 uGlnMetAsnAsnGlnGlyLysValAsnValThrValSerAsnThrSerM 1456
3629 ACACCAAAACACTACGCTTGGCAAGATTTCCGCGCTACGCGCAACAAAC 3678
1456 eTAsnGlnAsnTyLysSerSerSerGlnTyArgArgPheSerSerLysSer 1472

```

```

3679 GACCTGCGCAATCGGTATGCGAAGAAACCTGCGACGCG...CGCGT 3725
1473 ThrGlnThrGlnLeuGlnLysThrAspGlnThrLysSerAsnValGlnLe 1489
3726 CGGCATCTGCTTTTCCACACCGCGCAACAAACANTTTCGACGACGCA 3775
1489 uGlnGlyValPheThrValArgAsnSerAsnAsnPheAspLysAlas 1506
3776 TCGGCAACTCGCGCGCTTGGCCACGCGCGCTTTTCCGCAATCGCGC 3825
1506 eSerLysAsnThrLeuAlaGlnValAsnPheGlySerLysTyTyAla 1522
3826 ATCGGCGATTCGACATCGGCATCGACAGCGCGCGGTTTACGACGCG 3875
1523 AspAsnHisThrPtyrLeuGlyLysAspLeuGlyTyGlyLysPheGlnSe 1539
3876 CANTCTTTCAGACGCGATCGGAGGCAAAATCCGCGCGCGGTCTGAT 3925
1539 rAsnLeuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnP 1556
3926 ACGGCATTCAGCGACATACCGCGCGGTTTCCGCGATTCGCGATCGAA 3975
1556 heGlyLeuThrAlaGlyLysAlaPheAsnLeuLysAsnPheGlyLeuThr 1572
3976 CCGTACATCGCGCGCAACGCGCTATTTGTCGCAAAAAGCGATTACGCTA 4025
1573 ProLeuValGlyValArgTyLysSerTyLysSerAsnAlaAsnPheAla 1589
4026 CGAAACGTCATATGCGCACCGCGCGGTCTGCGTTCACCGTACCGNG 4075
1589 uAlaLysAspArgLysValAsnProLysSerValLysThrAlaPhe 1606
4076 CGGCATTAAGCAGATATTATTCATCAACCGCGCGACACATTCATC 4125
1606 InGlnValAspLeuSerTyThrTyHisLeuGlyLys...PheSerVal 1621
4126 ACNCTTATTTNAGCTGTCTAT...ACCGATCGCGCTTCCGCGCAAGT 4172
1622 ThrProLeuSerAlaArgTyAspThrAsnGlnLysSerLysI 1638
4173 CCGACACGCGGTCATACCGCGCTATTTGCTCGGATTCGCGCAACCC 4222
1638 e.....AsnValAsnGlnTyAspPheAlaTyAsnValGlnAsnGln 1653
4223 GCAGTCGCGAATGCGCGGTAAACCGCAAAATCAAGGTTTCACGCTGTC 4272
1653 InGlnTyAsnAlaGlyLeuLysLeuLysTyHisAsnValLysLeuSer 1669
4273 NTCCAGCTGCGCGCGCGCAAGNCCGCACTGGAACGCGAACAACGCGC 4322
1670 LeuLeuGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1686
4323 GGGCATCAATTAAGCTAC 4341
1686 aGlnLeuLysLeuSerPhe 1692

seq_name: p1r2:A41859

seq_documentation block:
1G4-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C:Species: Haemophilus influenzae
A:Variety: Strain HK715
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: A41859
J: Bacteriol. 174, 2913-2921, 1992
J: Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A:Reference number: A41859; M01D:92234949
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1702 <POU>

```


A:Cross-references: GB:M67489; NID:g148906; PIDN:AAA24966.1; PID:g148907
 A:Experimental source: strain HK715
 A>Note: sequence extracted from NCBI backbone (NCBIR:97282)
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

Alignment scores:
 Quality: 1517.00 Length: 1881
 Ratio: 1.651 Gaps: 53
 Percent Similarity: 48.857 Percent Identity: 24.615

Alignment block:

US-09-303-518d-651 x A41859

Align seg 1/1 to: A41859 from: 1 to: 1702

```

64 CGCTTCGCCCTGCTTACTTACCATATGCGCTGCTGCGCATTCCTTC 113
   ::::::::::: ::::::::::: :::::::::::
5 LysPheLysLeuAsnPhelLeuAlaLeuThrValAlaThrAlaLeuThrPr 21
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
114 CCAGCTTGCGGCGGACACACTTATTTGGCATCACTACCAATCTATAC 163
   | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
21 GyrThrGluAlaLeuValAlaArgAspAspValAspArgLnlPheA 38
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
164 GCGACTTGCGGCAAAATTAAGCAAGTTGCAAGTGGGCGGCAAGATAT 213
   ::::::::::: ::::::::::: :::::::::::
38 IAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
   ::::::::::: ::::::::::: :::::::::::
214 GAGGTNTCAACAAAAAGGGAGTGGTCCGCAATCAATGACAAAAAC 263
   ::::::::::: ::::::::::: :::::::::::
55 GluValAlaArgAspLysAsnAsnHisSerLeuGluAsnValLeuProAsnG 71
   ::::::::::: ::::::::::: :::::::::::
264 C...CGATGATGATTTTCTGCTGCTGC...CGTAACGGCGTGGCGG 307
   : ::::::::::: ::::::::::: :::::::::::
71 YlePrometIleAspPheSerValValAspValAspLysArgIleAla 88
   ::::::::::: ::::::::::: :::::::::::
308 CATGCTGGCGCATATATATGTCAGCGTGGCAAT...AAGCGCGC 354
   ::::::::::: ::::::::::: :::::::::::
88 hLeuIleAsnProGlnTyValAlaGlyValLysHisValSerAsnGly 104
   ::::::::::: ::::::::::: :::::::::::
355 TATAACAGCTGATTTTGGT.....GCGAAGCAAGAA 389
   ::::::::::: ::::::::::: :::::::::::
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlyAs 121
   ::::::::::: ::::::::::: :::::::::::
390 TCCCGATCAGACCGT.....TTTCTTACCAGATG 421
   || ::::::::::: :::::::::::
121 nAspLysSerHisArgAspValSerSerGluAsnArgTyrrPheSerV 138
   || ::::::::::: :::::::::::
422 TGAAGAAGATATATAT.....AAG 441
   || ::::::::::: :::::::::::
138 aGluLysAsnGluTyrrProThrLysLeuAsnGlyLysAlaValThrThr 154
   || ::::::::::: :::::::::::
442 CCTGACATTCACACCCCTTACACGGGATATACATATGCGGCTTGCA 491
   || ::::::::::: :::::::::::
155 GluAspGlnThrGlnLysAlaGlyArgLysPyrTyrrMetProAspGluAs 171
   || ::::::::::: :::::::::::
492 TAAATTTGTCACAGATCGAAGACCTGTGAA...ATGACGAGTGACATGA 538
   || ::::::::::: :::::::::::
171 PAspPheValThrGluValAlaProIleGluAlaSerThrAlaSerSera 188
   || ::::::::::: :::::::::::
539 GGGGGAATACCTATTCGATTAAGAAAAATATCCGAGCGTTCGCGATC 588
   ::::::::::: ::::::::::: :::::::::::
188 sPalagLysThrTyrrAsnAspGlnAsnLysTyrrProAlaPheValArgLeu 204
   ::::::::::: ::::::::::: :::::::::::
589 GGCTCAGGACACACACTATTTGGCGTTATGATGATGATCAACAACAGCGCAT 638
   || ::::::::::: :::::::::::
205 GlySerGlySerGlnPheIleTyrr.....LysLysGlyAsp.. 216
   || ::::::::::: :::::::::::
639 ATCTTACTCCGGCGCATGTTAATTTGGCGGCAATATACATATGAGGGTT 688
   ::::::::::: ::::::::::: :::::::::::
217 AsnTyrrSer.....LeuIleLeuAsnAsnHisGluValGly. 228

```

```

689 GGGGAATAT.....GGCTANTT 708
   || :::::::::::
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrrThrTyrrGlyIleAla 244
   :::::::::::
709 AGTTTGAGCGCGATGTGCGCCATGCCAACGAC.....TATGG 746
   || :::::::::::
245 GlyThrProTyrrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheCl 261
   :::::::::::
747 C.....C 748
   |
261 yAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnAsp 278
   |
749 CTATGCGGATTTGCGGTGGCGGACGCGGCTTGCCAAATGTTTAT 798
   || :::::::::::
278 rOleuThrAsnTyrrAlaValLeuGlyAspSerGlySerProLeuPheVal 294
   :::::::::::
799 TATGCAAAACAAACAAATTAATGCTCTGACAGGATTTTACAAACCG 848
   || :::::::::::
295 TyrrAspArgLysGlyLysTyrrPhePheGly.....Se 307
   :::::::::::
849 CTACCTTATTCGCGGCAAGGAAAGCTTTCCAGCTGATACGAAGAT 898
   || :::::::::::
307 rTyrrAspPheThrAlaGlyTyrrAsn.....LysLysSer 319
   :::::::::::
899 GGTTCCTACGATGACATTTTACAGAGCGATACATACCTCTTTTGA 948
   || :::::::::::
319 rPglngLntPrAsnIleTyrrLysProGlu..... 328
   :::::::::::
949 CCGCGCAGTAACGACATTTTCTTACATCCACAAACAGCGTACG... 996
   || :::::::::::
329 .....PheAlaLysThrValLeuAspLysAspThrAl 339
   :::::::::::
997 ..GGTACGTTACAGAAACCAACGAAAGTNTCC...AATCCAAACTTA 1042
   || :::::::::::
339 aGlySerLeuThrGlySerAsnThrGlnTyrrAsnTyrrAsnProThrGlyL 356
   :::::::::::
1043 AAGTACAGACAGTCCGACTGTTGACGAATCTTGAAT..... 1080
   || :::::::::::
356 yStrSerValIleSerAsnGlySerGluSerLeuAsnValAspLeuPhe 372
   :::::::::::
1081 .....GAACGTATTAAGAACGATTTACGCGGCGGAGGTGT 1118
   :::::::::::
373 AspSerSerGlnAspThrAspSerLys..... 381
   :::::::::::
1119 TAATCAGTACCCTCCAAAGCTTAACAACGCTGAAACCTTTCTTATG 1168
   || :::::::::::
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391
   :::::::::::
1169 ATTACGCAACGCGCAAACTCATCTATCAACAACGATCAACCAAGCGCG 1218
   || :::::::::::
392 ..ArgGlySerGlyThrLeuThrLeuAsnAsnHisLeuPglngLysAla 407
   || :::::::::::
1219 GCGGTTTGTATTTGAAGGTATTTACGGTC...TCGCGTGAACAA 1265
   || :::::::::::
408 GlyGlyLeuPhePheGluGlyAspTyrrGluValLysGlyThrSerAsp 424
   || :::::::::::
1266 CGAAAGCTGGCAAGCGCGGCTTCATATCAGTGAAGACAGTACCGTTA 1315
   || :::::::::::
424 rThrThrThrPylsGlyAlaGlyAlaSerValAlaAspGlyLysThrVal 441
   || :::::::::::
1316 CTGGAAGATTAACGCGGTGGCAACGACGCGCTGTCGAAATGCGGCA 1365
   || :::::::::::
441 hTrPrLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
   || :::::::::::
1366 GGCAGCTGCAGCTTAAGCAAGGGGAAACCAAGGCTGCATGACGCT 1415
   || :::::::::::
458 GlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySerLeuLysVa 474
   || :::::::::::
1416 GGGCGAGGTACAGTCATTTGGATTCAGAGGCAACGATTAAGCAAAA 1465
   || :::::::::::
474 IGlAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysV 491
   :::::::::::
1466 AACAAAGCTTTAGTAATCGCTTGNTCAGCGGCGGATCGGTGCA 1515

```



```
3094 CGTTACCACTCATCCGCAAGACGGAGTTCGGCTGCATATCG... 3141
:::|||||::: |||:::
974 LysTyrLeuArgAsnValAsnGlyArgTyrAspLeuTyrAsnProG1 990
3141 ..... 3141
990 uValGluLysArgAsnGlnThrValAspThrThrAsnIleThrThrProA 1007
3141 ..... 3141
1007 snAsnIleGlnAlaAspValProSerValProSerAsnAsnGluGluIle 1023
3141 ..... 3141
1024 AlaArgValGluThrProValProProProAlaProAlaThrProSerG1 1040
3141 ..... 3141
1040 uThrThrGluThrValAlaGluAsnSerLysGlnGluSerLysThrValG 1057
3141 ..... 3141
1057 LuLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnGlyGluVal 1073
3141 ..... 3141
1074 AlaGluGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGluVal 1090
3142 ..... GTCAAG 3148
1090 lAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGluIleLysG 1107
3149 AACAGAGCTTCCGACAACCTCGCAGCAGAGCCAAACAAACAG... 3195
|||::: |||:::
1107 LuThrAlaLysValGluLysGluGluLysAlaLysValGluLysGluGlu 1123
3195 ..... 3195
1124 LysAlaLysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSe 1140
3195 ..... 3195
1140 rGluThrSerProLysGlnAlaLysProAlaProLysGluValSerThrA 1157
3195 ..... 3195
1157 sPThrLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSer 1173
3195 ..... 3195
1174 ThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLysProAlaG1 1190
3196 ..... GCGGAAAAAGACACGCGCAAGCCTTGACCGCGTGA 3232
1190 uGluThrGlnProSerGluLysThrAsnAlaGluProValThrProValY 1207
3233 TTGGCGCGCGG... 3243
1207 alSerLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGlu 1223
3244 CGGCATGCCCGCCGAAACAGACAAAGCGTTCGCAACCGCGCGCGGCGC 3293
::: |||::: |||:::
1224 LysThrAlaLysValGluThrGluLysThrGlnGluProProGlnValAl 1240
3294 AGGC... 3297
1240 aSerGlnAlaSerProLysGlnGluGlnSerGluThrValGlnProGlnA 1257
3298 ..... GGGGAAATGTC...GGCATTTATGACGCGAGAGAGAG 3333
::: |||::: |||:::
1257 laValLeuGluSerGluAsnValProThrThrValAsnAsnAlaGluGluVal 1273

3334 AAAAACGGGTGACGGGATTAAGAC... 3360
::: |||::: |||:::
1274 GlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnPr 1290
3361 ..... AGCCNTTGGCGAAAC 3376
1290 cAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIleThrGluT 1307
3377 AGCGGAAAGCGGAACCGCGCGGNTAC... 3405
|||::: |||:::
1307 hrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaSerThrGlu 1323
3405 ..... 3405
1324 AspAlaSerGlnIleLysAlaAsnThrValAlaAlaAspAsnSerValAlaAs 1340
3406 ..... ACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGC 3445
::: |||::: |||:::
1340 nasSerGluSerSerGluProLysSerArg...ArgArgArgSerLies 1356
3446 CGCAACCGCAG... 3456
|||::: |||:::
1356 erGlnProGlnGluThrSerAlaGluGluThrThrAlaAlaSerThrAsp 1372
3457 ..... CCCCAACCGGAACCTCAACCCCAAC 3482
1373 GluThrThrIleAlaAspAsnSerLysArgSerLysProAsnArgArgSe 1389
3483 GCAG... 3486
1389 rArgArgSerValArgSerGluProThrValThrAsnGlySerAspArgS 1406
3487 ..... CCGGACCTGATNAGCGTTATGCCAATACGGTTTG 3522
|||::: |||::: |||:::
1406 erThrValAlaLeuArgAspLeuThrSerThrAsnAlaValAlle 1422
3523 AGTGAATTTCCGCC... 3557
|||::: |||::: |||:::
1423 SerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLysAl 1439
3558 CGTACAGAGCAATTTGGACCGCGTG...TTTGGCAAGACCGCGCAACG 3604
|||::: |||::: |||:::
1439 aValSerGlnIleIleSerGlnLeuGluMetAsnGlnGluGlyLysAla 1456
3605 CNGTTTGGACAAGCNGCATCCGGNACACCAACACTACCTTGCGACAGAT 3654
|||::: |||::: |||:::
1456 snValITrPValSerAsnThrSerMetAsnGluAsnTyrSerSerSerGln 1472
3655 TTCCGCGCTTACCGCCCAACACCGACCTGGCGCAATGCGTTCGAGAA 3704
|||::: |||::: |||:::
1473 TyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTyrAspG1 1489
3705 AAACCTCGGACGCGG...CGCGTGGCATCGTGTTCGACACACCGGA 3751
|||::: |||::: |||:::
1489 nThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrValArgA 1506
3752 CGGAAAAACNCTTCGACGACGCGCATCGGCACTCGGACGCGGTGCCAC 3801
::: |||::: |||:::
1506 snSerAsnAsnPheAspLysAlaSerSerLysAsnThrIleuAlaGlnVal 1522
3802 GCGCGCGTTCGCGCAATACGCGATCGGACGAGTTCGATCGGCATCGAG 3851
::: |||::: |||:::
1523 AsnPheTyrSerLysTyrTrpAlaAspAsnHisTrpTyrLeuGlyLeuAs 1539
3852 CACGGGCGGCGGTTTACGACGCGGANTTCACAGCGGATCGAGGCA 3901
::: |||::: |||:::
1539 pleuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHisAsnAlaL 1556
3902 AAATCCGCGCGCGCGTGCATTCAGGCATTCAGGCAATCAGAGATACCGCGC 3951
|||::: |||::: |||:::
1556 yPheAlaArgHisTrnAlaGlnPheGlyLeuThrAlaGlyLysAlaPhe 1572
3952 GGTTCGGCGGATTCGCGCATCGAACCGTTCATCGGCGCAACCGCTATT 4001
```



```
782 GTTGGCCAAATGTTATTTATGACAAAACAATAAATGCTGCTCAAC 831
|||||
279 LysSerProLeuPheAlaPheAspLysGlnLysAsnGlnThrValPheLeu 295
832 GGAGTTTACAAACCGGCTACCTTATTCGGCAGGAAACGGTTTCCA 881
|||||
296 GltThrLysPThrLysPThrLysGly..... 305
882 GGTGATACGCAAGATTGGTTTACGATGCAATTACAGGGCAT.... 927
|||||
306 .....LysLysSerThrGlnLysThrPheAsnLysLysGlnPheA 320
928 .....ACACATACCGTCTNTTTGACCCGGCAGATACGGA 963
320 LAspLysLysLysGlnHisAspAsnAlaGlyThrValLysGlyAsnGly 336
964 .....CATTTTCTTTACATCCAAACAACACGAGTACGGTTAAC 1007
|||||
337 GlnHisHisThrLysThrThrGlyThrAsnSerHisIleGlySerThr... 352
1008 AGAACCACGAAAGGTMTCCATCCAAAGCTTAAAGTACAGACAGTCC 1057
|||||
353 .....AlaValA 355
1058 GACTGTTTACGCAATCTTGAATGAATGATAAAGAACAGTTACGCG 1107
|||||
355 rGluValAsn.....AsnGluGlyAsp..... 362
1108 GCAGGGGGGTATATCATGACGCTCCAGGTTAAACACGAGTAAACCT 1157
|||||
363 .....AlaAsnAsnGlnGlnHisnVa 369
1158 TTCTTTATGATTAAGCGACGCAACTCATCTTATCAACAACATCA 1207
|||||
369 LThrPheGluAsp.....AsnGlyThrLeuValLysAsnGlnHisIleA 384
1208 ACCAAGCGCGCGGCTTTGATTTGAAGTGATTTTACGGTCTGGCT 1257
|||||
384 snGlnGlyLysGlyLeuPhePheLysGlyAspLysThrValLysGly 400
1258 GAAACACAGCAA...ACGTGCGAAGCGCGGCTTCATATCATGACGA 1304
|||||
401 AlaAsnAsnAspLysThrLysLeuLysAlaGlyLysPheValAlaAspG 417
1305 CAGTACCGTTACTGAAAAGTAAACGGCGGTGCAACGACCGCTGTCCA 1354
|||||
417 LysLysValValLysGlnValLysAsnProAsnGlnLysPheLysAla 434
1355 AAATGCGCAAAAGCGACGCTGCAAGCTTCAACCAAGGGAACCAAGC 1404
|||||
434 LysLysGlyLysGlyThrLeuGlnLysAsnGlyThrGlyValAsnGln 450
1405 TCGATCAGCGTGGCGACGCTGACATCATTTTGGATCAGCAGGACGA 1454
|||||
451 GlnLeuLysValGlyAspGlyThrValLysLeuAsnGlnLysAlaAsp 467
1455 TAAAGCGCAAAAACAAGCTTGTGAATCGCTGTCAGCGGACAGG 1504
|||||
467 aAspLysLysValGlnAlaPheSerGlnValGlyLysValSerGlyArg 484
1505 GTACGTCGCACTGAATGCGGATATACGTTCAACCCGACAACTCAT 1554
|||||
484 LysThrLeuValLysAsnSerSerAsnGlnLysProAspAsnLeuLys 500
1555 TTGGGCTTTCGGCGGACGTTTGGATTTAAAGGCGATTCGTTTGGTT 1604
|||||
501 PheGlyPheArgGlyGlyArgLeuAspLysAsnGlnLysAsnLeuThrP 517
1605 CCACCGTTTCAAAATACCGATGAAGGCGGATATGMCNATCATATG 1654
|||||
517 eGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAsn 534
```

```
1655 CCACAACAATCCACCGTTTACGATTAAGGGAATGAATTAACAA 1704
|||||
534 hAspHisLysSerThrLysThrLeuThrGlyLysSerLeuLysThrAsn 550
1705 CCGAGT..... 1710
551 ProAsnSerLeuSerValHisSerIleGlnAsnAspLysAspGluAsp 567
1711 .....GTAAGATA 1720
567 PThrSerThrLysThrArgProArgArgProLysProGlnGlyAspL 584
1721 TC.....AATGACTT 1731
584 eThrLysLysAsnLysThrArgLysThrAlaLeuLysSerGlyArgLeu 600
1732 AAT..... 1734
601 AsnAlaProMetProGluAsnGlyValAlaGlnAsnAsnAspThrLys 617
1735 .....TACAGCAAGAAATTGCC..... 1752
617 eMetGlyThrThrGlnGlnGlnValAlaArgLysAsnAlaMetAsnHis 634
1753 .....TACAACGTTGTTTGGCGAGAAAGATACG 1782
634 snAsnArgArgIleGlyAspPheGlyGlyPhePheAspGlnGluAsnG 650
1783 ACCAAAACGACGCGCGCTCAACCTTGTATTCAGCCGCGCGCAGAGA 1832
|||||
651 LysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlnLysSerAla 667
1833 CCGCACCCGCTGCTTCCGCGGACAAATTTAAACGCAATCACGC 1882
|||||
667 nLysArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysIleSer 684
1883 AAACAAGCGCAACTGTTTTCAGCGGACGACCGCAACCGCAGCTAC 1932
|||||
684 aThrGlnGlnLysValLysLeuLeuSerGlyArgProThrProHisAla 700
1933 AATCATTTAGGAACGGGCTGTCAAAATGCAAGT.....ATCCACA 1976
|||||
701 AspPheValAsnLysSerSerAlaArgLysAspLysHisPheSerLys 717
1977 AGGAGAATGCTGTGGGACACGCTGATGTCMACCGACGTTTAAACGG 2026
|||||
717 masGlnValValPheGlnAspAspThrLysAsnArgThrPheLysAla 734
2027 AAAATTCATATTT...CAGGCGGCGACGCGGTGATTCGCGCATGTT 2073
|||||
734 LysGlnLysValLysAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2074 GCCAAATGCAAGCGCATTTGNCATTTGAGCAATTCAGCCGACAGTTT 2123
|||||
751 SerAspLysThrAlaAsnLysThrAlaThrAspAsnAlaLysValAsn 767
2124 TGGGTGCGCACCGCATCAAGCCATACATTCGTACACGTTGCGACTGGA 2173
|||||
767 uGlyLysLysLysAsnLysAspGlnLysValLysSerAspLysThr 782
2174 CNGGTGACAAATTTGTTCGAANAANAATTAACGACGATAAGTGATT 2223
|||||
782 hArgLysValThrCys..... 787
2224 GCTTCATGTACTAAGACGACGACGACGACGACGACGACGACGACGAC 2273
|||||
787 ..... 787
2274 CGTNTNTTAAANCCTCNCGCGCTGCTCNCACGCTTAAAGCAATCTAGT 2323
|||||
788 .....AsnThrGlyAsnLeuSer... 793
2324 CAATGCGCATACGTTATACAGTCAGCCACAGACCGCACCCAA...AAC 2370
```

```
794 .....AspLysAlaLeuAsnSerPheAspAlaThrArgIleAsn 806
2371 GGCAACCTTAGCCCTGTGGCAATGCCCAAGCAACATTATTAAGCCAC 2420
807 GlyAsnValAsnLeuAsnGlnAsnIleAlaLeuValLeuGlyLysAlaI 823
2421 ATTAAAGGCAACATCGGNTTGGCAATGCTTCATTATTAATCAAGCA 2470
823 aleuTrpGlyLysIleGlnGlnGlnLysSerArgValSerLeuAsnG 840
2471 ACAAGCCGCGCAAAAGCGAGTCTGACGCTTCCGACAAGCTAAGCA 2520
840 LnhIS..... 841
2521 AACGTAAACCATTCGCACTCAACGGCAATGTCCTAGCCGATAAGC 2570
841 ..... 841
2571 AGTATTCATTTTGAAACAGCCGCTTACCGGACACTAGCGGACGA 2620
841 ..... 841
2621 AGAAGACAGCATTAACCTTAAAGACAGCAATGAGACGTCGCGTACGC 2670
842 .....SerLysThrPheLeuThrGlyAsp 849
2671 ACGGAATTAGGCATTTAAACCTTGACAGCCACCATTAACATCAATTC 2720
850 SerGlnValHisAsnLeuSerIleAlaSerHisIleHisLeuAsnAs 866
2721 CGCCTATGCGCAGATGCTGCAGAGCGGCAACCGGACGNGTGCAGACA 2770
866 nAla.....SerAspAlaGlnSerAla..... 873
2771 CGCCGCGCGCGCTTCGCGCGCTTCCTATTATCCGTTACACGCGCACT 2820
873 ..... 873
2821 TCGGTAGATCCCGTTTCAACAGCTGACGGTAAACGGCAATTAAGACNG 2870
874 .....AsnLysTyrHisThrIleLysIleAsn...HisLeuSerG1 886
2871 TCAAGGACATTCGCGTTTATGTCGGAACCTTTCGGTACCGAAGCGACA 2920
886 yAsnGlnHisPheHisTyrLeuThrAspLeuAlaLysAsnLeuGlyAspL 903
2921 AATTGAAGCTGGCGAAAGTTCCGAAGNACTTACACCTTGCGGCTCAAC 2970
903 yValLeuValLysGlnSerAlaSerGlyHisTyrGlnLeuHisValGln 919
2971 AATTACCGGCAACGAAACC.....GTAAACCTCGATCAATTGACGGTACT 3014
920 AsnLysThrGlyLysProAsnGlnLysLeuAspLeuPheAspAla 936
3015 GGAAGGAGAAAGACAAACACCGCTGCGAAACCTTAATTTCACCCCTGC 3064
936 rSerValGlnAspArg.....SerArgLeuPheValSerLeu 949
3065 AAAAGCAACAGCTGATCGCGCGCGGTGCGTACCAACTCATCCGCAAA 3114
949 lAsnHisTyrValAspLeuGlyAlaLeuArgTyrThrIleLysThrGln 965
3115 GACGGGAGATTCCGCTGCATTAAT.....CCGGT 3143
966 AsnGlyIleThrArgLeuTyrAsnProTyrAlaGlyAsnGlyArgPro 982
3144 CAAGAAGACAAGAGCTTCCGACAACTCGCAAGGAGGAGAGGCAAAAC 3193
982 lLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLys 999
3194 AGCGGAAAAAGACAACGCG..... 3213
999 lathrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValAl 1015
3214 .....CAAGCCCTGACGGCTGATGCGCG 3239
1016 AlaProProSerProGlnAlaAsnGlnAlaGlnIleAlaLeuAlaGlnGlnI 1032
3240 CGGCGCGATGCGCGCGGAAAGACAAAGCGTTCGCAACGCGCGCGC 3289
1032 nAlaLysAlaGlnGlnValLysArgGlnGlnIleAlaGlnAlaGlnLysV 1049
3290 NCGCAGCGCGGAAAGTGTGCGCATTAATGACGCGGAGGAGGAGGAAAAA 3339
1049 aAla.....ArgGlnLysAspGlnGlnLys 1058
3340 CGGCTGACAGCGGAT..... 3354
1059 ArgLysAlaAlaGlnIleAlaArgGlnGlnGlnIleAlaArgLysAlaAl 1075
3355 .....AAGACAGCGCNT 3367
1075 aGlnLeuAlaAlaLysGlnLysAlaGlnIleAlaGlnLysAlaArgLysI 1092
3368 TGGCGAAACAGCGCGGAAAGCGGAAACCGCGGNTACACGCGCTTCCCG 3417
1092 euAlaArgGlnLysAlaGlnIleAlaSerHisGlnAlaAsnAlaLysPro 1108
3418 CGGCGCGCGGCGCGCGCGGATTTGCCGCAACCG..... 3453
1109 LysArgArgArgArgArgAlaIleLeuProLysProProAlaProValP 1125
3453 ..... 3453
1125 eSerLeuAspAspTyrAspAlaLysAspAsnSerGlnSerIleGlyA 1142
3453 ..... 3453
1142 snLeuAlaArgValIleProArgMetGlyArgGlnLeuIleAsnAspTyr 1158
3453 ..... 3453
1159 GlnGlnIleProLeuGlnGlnLysAspGlnAlaGlnGlnLysArg 1175
3453 ..... 3453
1175 gGlnAlaThrGlnPheHisSerLysSerArgAsnArgArgAlaIleSers 1192
3453 ..... 3453
1192 eArgLysProSerSerAspGlnAspAlaSerGlnSerValSerThrSerAsp 1208
3453 ..... 3453
1209 LysHisProGlnAspAsnThrGlnLeuHisGlnLysValGlnIleThrAlaGln 1225
3454 .....CAGCCC.....CAGCCGACACTCAACCCCAACCCGACGCG 3490
1225 yLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAlaAlaAlaAla 1242
3491 ACCTGATNAGCCGTTATGCCAATAGCGGTTTGATGATTTTCGCGACG 3540
1242 spAlaValSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSer 1258
3541 CTCACAGCGCTT.....TTGCGCGCTACAGACGAGATTT 3572
1259 ThrGlnSerIleLeuLeuAsnThrGlnGlyAlaTyrLeuThrArgHisIleAl 1275
3573 GGAACGCGGTGTTTCCGAAGACGCGGCAACGCGGTTTGACAGACGCA 3622
1275 aGlnLysSerArgAlaAspAlaGlnLysAsnSerValTyrPheSerAsn 1292
3623 TCCGNAACCAACACTACCGGTCGCAAGATTTCCGCGCTACCGGCA 3672
1292 hrGlyTyrGlyArgAspTyrAlaSerAlaGlnIleTyrArgLysPheSer 1308
```

```

3673 CAACCGGACCTGGCCAAATCGGATGTCAGAAAAAACCCTGGCAGCCGG... 3720
      :: |||||||.....::|||.....::|||.....::|||.....
1309 LysArgThrGlnThrGlnIleGlyIleAspArgSerLeuSerGluAsnMe 1325
      ::|||.....::|||.....::|||.....::|||.....
3721 .CGCGTCGGCATCCTGTTTTCGCAACAACCGCAGCAAAACATCGTCAGC 3769
      ::|||.....::|||.....::|||.....::|||.....
1325 tGlnIleGlyIleValLeuThrThrSerAspSerGlnHisThrPheaspG 1342
      ::|||.....::|||.....::|||.....::|||.....
3770 ACGGCATCGGCAACTCGGACGCGCTTGCACGCGCGCGCTTTCGGGCA 3819
      ::|||.....::|||.....::|||.....::|||.....
1342 InAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeuThrGlyLys 1357
      ::|||.....::|||.....::|||.....::|||.....
3820 TACGGCATCGGC...AGGTCGACATCGGCATCGGAGCAAAATCGCGCGGTTT 3866
      ::|||.....::|||.....::|||.....::|||.....
1358 TyrThrLeuAsnAspAlaThrPyrValAlaGlyAspIleValIleGlySe 1374
      ::|||.....::|||.....::|||.....::|||.....
3867 TACGACGGCATCTCTTCAGAGCGCATCGGAGCAAAATCGCGCGCGG 3916
      ::|||.....::|||.....::|||.....::|||.....
1374 rLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgThrS 1391
      ::|||.....::|||.....::|||.....::|||.....
3917 TCCTGATTCAGGATTCAGGACGATACCGCGCGCTTTCGGCGGATTC 3966
      ::|||.....::|||.....::|||.....::|||.....
1391 erIleGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPhe 1407
      ::|||.....::|||.....::|||.....::|||.....
3967 GGCATCGAAACCGTACATCGCGCGCAACGCGCTATTGTCCAAAAAGCGGA 4016
      ::|||.....::|||.....::|||.....::|||.....
1408 GlnIleValProSerAlaGlyIleArgThrSerArgLeuSerSerAlaAs 1424
      ::|||.....::|||.....::|||.....::|||.....
4017 TTCGCCCTCAGAAAACGTCATATCCGCAACCCCGGCTTCGCTTCACCC 4066
      ::|||.....::|||.....::|||.....::|||.....
1424 PyrThrLysLeuGlyAspAspSerValLysValSerSerMetAlaValIle 1441
      ::|||.....::|||.....::|||.....::|||.....
4067 GNTACCGCGCGGCGATTAAGCGAGATTATTCATTAACCGCGCGCAACAC 4116
      ::|||.....::|||.....::|||.....::|||.....
1441 hrLeuThrAlaGlyLeuAspPheAlaThrArgPheLys...ValGlyAsn 1456
      ::|||.....::|||.....::|||.....::|||.....
4117 ATATTCATCAACNCTTATTATTCAGCGCTTCATACCGATCGCGCTTCGGG 4166
      ::|||.....::|||.....::|||.....::|||.....
1457 LeuThrValLysProLeuLeuSerAlaAlaThrPhe...AlaAsnTyrC1 1472
      ::|||.....::|||.....::|||.....::|||.....
4167 CAAGTCGCAACGCGGTCATACCGCNGTATGCTCAGGATTCGCGCA 4216
      ::|||.....::|||.....::|||.....::|||.....
1472 LysGlyGlyValAlaAsnValGlyGlyLysSerPheAlaThrLysAlaAsp 1489
      ::|||.....::|||.....::|||.....::|||.....
4217 AAACCGCGAGTCGGGATGCGGCTAAACCGCAATCAAGTTCACG 4266
      ::|||.....::|||.....::|||.....::|||.....
1489 snGlnGlnIleThrSerAlaGlyValAlaLeuLeuThrArgAsnValThr 1505
      ::|||.....::|||.....::|||.....::|||.....
4267 CGTCTCCATCCAGCTGCGCGCGCAAGAGCCGCAACTGGAGCGCAACA 4316
      ::|||.....::|||.....::|||.....::|||.....
1506 LeuAsnValAlaAsnGlySerIleThrLysGlyLysGlnLeuGlnLysGln 1522
      ::|||.....::|||.....::|||.....::|||.....
4317 CAGCGCGGCGATCAATTAAGCTACCGCGG 4347
      ::|||.....::|||.....::|||.....::|||.....
1522 sSerGlyGlnIleLysIleGlnIleArgPhe 1532

```

seq_name: pir2:C41859

seq_documentation_block:

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae

C:Species: Haemophilus influenzae

A:Variety: strain HK613

C:Date: 04-Mar-1993

#sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: C41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty

A:Reference number: A41859; M01D:92234949

A:Accession: C41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613
A:Note: sequence extracted from NCBI backbone (NCBI:P:97285)
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 1475.00 Length: 1924
Ratio: 1.517 Gaps: 64
Percent Similarity: 50.520 Percent Identity: 25.364

alignment_block:

US-09-303-518d-651 x C41859 ..

Align seg 1/1 to: C41859 from: 1 to: 1849

```

64 GCGTCTGCGCGCTTACTTACCATATGCTGTCGCTTCGCTTCCTTC 113
      ::|||.....::|||.....::|||.....::|||.....
5 LysPheLysLeuAsnPheIleLeuThrValAlaThrValAlaLeuThrPr 21
      ::|||.....::|||.....::|||.....::|||.....
114 CCAGCTTGGCGCGGACACACTTATTTGGCATCACTACATCAATATCATC 163
      ::|||.....::|||.....::|||.....::|||.....
21 CThrThrGlnAlaAlaLeuValArgAspAspValAspThrGlnIlePhe 38
      ::|||.....::|||.....::|||.....::|||.....
164 GCGACTTTGGCGGAAATATAAGCAAGTTTGCAGTCGGGCGCAAGATATT 213
      ::|||.....::|||.....::|||.....::|||.....
38 rGAspPheAlaGlnLysAsnGlyLysPheSerValGlyAlaThrAsnVal 54
      ::|||.....::|||.....::|||.....::|||.....
214 GAGGTNTACAAAACAAAAGGAGTGGTGGCAATCAATCAATCAAAAAC 263
      ::|||.....::|||.....::|||.....::|||.....
55 GluValAlaGAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
      ::|||.....::|||.....::|||.....::|||.....
264 C...CGATGATGATATTTTCTGTGTGTCG...CGTAACGCGCGTGGCG 307
      ::|||.....::|||.....::|||.....::|||.....
71 YlePrometIleAspPheSerValAlaAspValAlaAspLysArgThrLe 88
      ::|||.....::|||.....::|||.....::|||.....
308 CATTCGTCGGGATCAATATTTGTGAGCCGTGCGACAT...AACGGCGGC 354
      ::|||.....::|||.....::|||.....::|||.....
88 hrLeuValAsnProGlnThrValAlaGlyValLysHisValSerAsnGly 104
      ::|||.....::|||.....::|||.....::|||.....
355 TATAACAAGCTGATTTTGTGT...CGCGAAGGAAGNAA 389
      ::|||.....::|||.....::|||.....::|||.....
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
      ::|||.....::|||.....::|||.....::|||.....
390 TCCCGATCAGCACCGT.....TTTTCTTACCAATTG 421
      ::|||.....::|||.....::|||.....::|||.....
121 nAlaLysSerHisArgAspValSerSerGlnGluAsnArgThrThrV 138
      ::|||.....::|||.....::|||.....::|||.....
422 TGAAGAAGATATTAATTAAGCTGACAA... 450
      ::|||.....::|||.....::|||.....::|||.....
138 alGluLysAsnAsnPheProThrGlnAsnValThrSerPheThrLys 154
      ::|||.....::|||.....::|||.....::|||.....
451 .....TCACACCTTACACAGCGGATTAACATTCGCGCGTTC 488
      ::|||.....::|||.....::|||.....::|||.....
155 GluGluGlnAspAlaGlnLysArgArgGluAspThrThrMetProArgLe 171
      ::|||.....::|||.....::|||.....::|||.....
489 GCATAAATTTGTCCAGATGACAGACTGTTCGA...ATGACGAGTGACA 535
      ::|||.....::|||.....::|||.....::|||.....
171 uAspLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaAsn 188
      ::|||.....::|||.....::|||.....::|||.....
536 TGAAGGGGAATACCTATTCGATTAAGAAAATATCCGCGCGTTCGCG 585
      ::|||.....::|||.....::|||.....::|||.....
188 snAsnLysGlyGluThrAsnAsnSerAspLysThrProAlaPheValArg 204
      ::|||.....::|||.....::|||.....::|||.....
586 ATCGGCTCAGGACACCATAT.....TGGCGTTAT..... 615
      ::|||.....::|||.....::|||.....::|||.....
205 LeuGlySerGlySerGlnPheIleThrLysGlySerArgThrGlnLeu 221
      ::|||.....::|||.....::|||.....::|||.....
616 .....GATGATGACAAACACGCGCATTTATCTACTCGCGCGCAT 655
      ::|||.....::|||.....::|||.....::|||.....
221 uIleLeuThrGlnLysAspLysGlnLysLeuLeuArgAsn.....T 236
      ::|||.....::|||.....::|||.....::|||.....
656 GGTAAATGGCGCAATACATATCGAGGTTGGGGAATATATGCGTA 705

```



```
11 .....:
236 rpaervaliglylasp...Asnleuglueuvaliglyasnalatythr 251
706 NTAGTTGAGGCGC.....GATGGGCGCATGCAACGAC..... 741
252 TyrlilylealaglythrProTyllysvalasnhisgluasnsnglyle 268
742 .....TATGGC..... 747
268 uileglypnecllyasnseryslugluhisseraspprolysglylel 285
748 .....CCTATGCCGATTGCGAGTGGCGAGCGGACGCGTTGC 786
285 euserglinspproleuthrasnlyralavalieuglyaspserglyser 301
787 CCATGCTTATTATGCAAAACAACAATTAATGCTGCTCAACGAGT 836
302 Proleuphevaltyrasparygludysglystirpleupheugly.. 317
837 TTACAAACCGCGTACCTTATTCGGCGAGGAAACGGTTTCAGCTGA 886
318 .....SerTyraspphetrpalaglytyrasn..... 326
887 TACGCAAGATTGGTCTTACGATGACATTTACAGAGCGGATACATACC 936
327 ..Lyslysertrpdinglutrpasnlletyrllyshisgluphealaglu 342
937 GCTTTTGTGAACCGCGCAGTACGACATTTTCCTTATCATCCACAA 986
343 LyslletyrlnglntyrSerlalely.....Serleuthnglyseras 357
987 C.....AACGTTACGGGTTACGGTAAACAGAA 1012
357 nthrclntyrthrtrpgrlnalathrglyserthrillerthrlyg 374
1013 CCACGAAAGGCTNTCCATCCAAAGCTTAAAGTACACAGCTCGACTG 1062
374 lyglgluproleu..... 378
1063 TTGACGATCTTTGATGAATGAATGATTAAGAACCACTTACGGCGCAG 1112
379 .....Servalasprouthraspoly..... 385
1113 GGGTGTATACGATACCGTCCAGGTTAAACACGCTGAACACCTTTCTT 1162
386 .....Lysasplysproasnhisgllysserlletnrl 397
1163 TTATGATTACGCAACGCAACATCTTATCAACAACATCAACCA 1212
397 eu.....Lysglyserglythrleuthrleuasnsnhisllenspdl 411
1213 GCGCGGGGGGTTGATTATTTGAAGTGAATTTACGCTC...TCGCCGTA 1259
412 gllyalaglyglyleuphehegluglyasptrygluvallysglythrse 428
1260 AAACAACGAAACGTGGCAAGCGCGGCTCATATCACTAGACAGCAGTA 1309
428 raspserrtrthrtrptrygllyalaglyalaservalalaaaspglylvt 445
1310 CGGTACTGTGAAGTAACGAGCGGTGGCAACGCGCTGCCAAATC 1359
445 hrvalthrtrptryvalhisasnprolystyrasptrygluvallystille 461
1360 GCGAAGGACGCTGCACGTTCAAGCAAGGAGGAAACCAAGGCTCGAT 1409
462 glylysglythrleuvalvalaglylysglylysasnglyleu 478
1410 CACCGTGGCGACGCTACAGTCAATTTGATCAGCAGGACAGCATAAAG 1459
478 ulysvaliglyaspglythrvalilleuulysglinsalaaaspalaasna 495
1460 GCACAAACACGCTTACTGAATCGGCTGTGTCAGCGGCGGCTAG 1509
:::
495 snlyvalglinalphesergluvalglyllevalserglyargserthr 511
1510 GTGCAACTGATACCGGATATACATTCAACCCGCAACACTCTATTGG 1559
512 Leuvalleuasnsaspasyglvalaspasnserrlletyrlphegl 528
1560 CTTTGGCGGCGAGCTTGAATTAACGGGATTCGTTGCTGCCACC 1609
528 ypheargglyglythrleuaspleuasnsnlyasnserrleuthrphesapn 545
1610 GATTAACAAATACCGATGAAGGCGGATGATTCGATCATATATCCACA 1659
545 istleargsnlleaspspoglyalargvalalasnhisasnmetthr 561
1660 ACACATCCACCGCTTACATTCACAGGAATGAAGATATTACACACCG.. 1707
562 AsnThrSerAsnlethrillerthrlygluserrleuillerhrasproas 578
1707 ..... 1707
578 nthrillerSerTyrasnilleglualaclinsaspasphsiproleua 595
1707 ..... 1707
595 rglleargserlleprotyrarggluleuTyrrheasnglnaspnary 611
1707 ..... 1707
612 SerTyrrTythrleuulysglyalaserthrArgSergluleuprogI 628
1708 ....AGTGTAAGAAT..... 1719
628 nasnseryglyluserasngluasntrpleuTyrmelglyargthrsera 645
1720 .....ATCAATAGACTTAATTACACCAAGCAAT 1749
645 spglualalysargasnvalmelasnhisilleasnsngluargmetasn 661
1750 GCGTACACGCGTTGTTGGCGAGAAAGATACGACCAACAAAG...AACGG 1796
662 glypheasnglytyrrpneglygluglugluthrlyalathrglinsncl 678
1797 GCGGCTAACCTGTATTACCGCGCCGCGAGAAAGCGGACCCGCTGC 1846
678 ylyslleuasnvalthrphesasnlylysaseraspglinasargpheleu 695
1847 TTTCCGGCGGAACAATTTAAACGCAACATCCACCAACAAACGCAAA 1896
695 euthrglyglythrAsnleuasnglyAspleuasnvalglulysglythr 711
1897 CTGTTTTCAGCGGACGACGACGACGACGCTACAAATCATTTAGGAAG 1946
712 leupheleuserglyarprothrprohnsalargaspillealaglyl 728
1947 CGGCTGTCAAAATGAGAGT.....ATCCCAAGAGGAATCGTGT 1990
728 eserSerThrlyslasprohnsphethrgluasnsnglivalvaly 745
1991 GCGACAAACGATGATCNACCGCACGTTTAAAGCGGAAATTTCCATAT 2040
745 algluaspsprpilleasnarghasnbheylsalatrtrhrmetasval 761
2041 CAGGCGCGGCGGCGGATTC...CGCAATGTCCAAGATGCAAG 2087
762 ThrlylinalaserleuTyrrserglyrghasnvalalasnlethrse 778
2088 CGATTGMCATTTGACATACCGCCCAAGCAATTTTGTGTGCGCACCGC 2137
778 rasnlllethrleasernsasnalaglnvalhisilleglytyr.....L 793
2138 ATCAAGCATATACATGCTACACGTTTCGACACTGACAGCTGCACAAAT 2187
793 ystnnglyaspthrvalcysvalargserasprtyrthrnglyTyrrvalthr 809
```



```

1377 GlnAlaIysProGlnThrGluProAlaArgGluAsnValSerThrValAs 1393
3501 ..... 3501
1393 mThrLysGluProGlnSerGlnThrSerAlaThrValSerThrGluGlnP 1410
3501 ..... 3501
1410 roAlaLysGluThrSerSerAsnValGluGlnProAlaProGluAsnSer 1426
3501 ..... 3501
1427 IleAsnThrGlySerAlaThrThrMetThrGluThrAlaGluLysSerAs 1443
3501 ..... 3501
1443 PLYSPROGlnMetGluThrValThrGluAsnAspArgGlnProGluAlaA 1460
3501 ..... 3501
1460 snThrValAlaAspAsnSerValAlaAsnAsnSerGluSerSerGluSer 1476
3502 ..... CGTTATGCCAATACCGCGTTGAGTGAATTTCCGCCAC 3539
1477 LysSerArgArgArgArgSerValSerGlnProLysGluThrSerAlaG 1493
3540 GGTCAACAGCGCTTTCCGCCGTACAGGAC..... 3567
1493 uGluThrThrValAlaSerThrGlnGluThrThrValAlaAspAsnSerValS 1510
3567 ..... 3567
1510 eThrProLysProArgSerArgArgThrArgSerValGlnThrAsn 1526
3568 ..... GAATGGACCGCGCTTTCCGCCAGACGGCGGCA 3602
1527 SerTyrGluProValGlnLeuProThrGluAsnAlaGluAsnAlaGluAs 1543
3603 CGCN..... 3606
1543 nValGlnSerGlyAsnAsnValAlaAsnSerGlnProAlaLeuArgAsnL 1560
3606 ..... 3606
1560 eThrSerLysAsnThrAsnAlaValLeuSerAsnAlaValAlaLysAla 1576
3606 ..... 3606
1577 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisTleSerG 1593
3607 ..... GTTGGACAGCAGCNGCATCC 3625
1593 nLeuGluMetAsnAsnGlnGlyGlnTyrAsnValTrrPleSerAsnThrS 1610
3626 GGNACACCAACACTACCGTCCGAAGATTCCGCCCTACCGCCACAA 3675
1610 eMetAsnLysAsnTyrSerSerGlnGlnTyrArgArgPheSerSerLys 1626
3676 ACCGACCTGGCCAAATCGGTATGCAAAACCTGGCAGGCGG...CG 3722
1627 SerThrGlnThrGlnLeuGlyTrrPaspGlnThrIleSerAsnValG 1643
3723 CGTGGCAGATCTGTTTCGCACACACGAGCAGAAACATCTTCGACGAG 3772
1643 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPhaAspLysA 1660
3773 GCATCGCAACTCGCACGGCTTCCACAGCGCGCTTTCCGGGCAATAC 3822
1660 LaserLysAsnThrLeuAlaGlnValAlaAsnPhaTyrSerLysTyrTyr 1676
3823 GGCATCGGCGAGTTCCGATCGCATCGCATCGAGCGCGGCTTTAGCAG 3872
::: ..... III III ::

```

```

1677 AlaAspAsnHisThrTyrLysGlyIleAspLeuGlyTyrGlyLysPheG 1693
3873 CGGCANTCTNTTCAGACGCGATCGGAGCAAAATCCGCCGCCGCTCTCC 3922
1693 nSerAsnLeuGlnThrAsnAsnAsnAlaLysPheAlaArgHisThrAlaG 1710
3923 ATTACGGCATTCAGCAGCATACCGCGGCTTCCGGCGATTCGGCAGTC 3972
1710 InIleGlyLeuThrAlaGlyLysAlaPheAsnLeuGlnLysnPheAlaVal 1726
3973 GAACCGTACATCGCGCAACCGCTATTTCTGTCAAAAGCGGATTACCG 4022
1727 LysProThrValGlyValAlaArgTyrSerTyrLeuSerAsnAlaAspPheAl 1743
4023 CTACGAAAACGTCAATATCGCCACCCCGCTCTTCGCTTCACCGTACG 4072
1743 aLeuAlaGlnAspArgIleLysValAsnProIleSerValLysThrAlaP 1760
4073 GNGCGGCATTAAAGCAGATTATTCATTCAAACCGCGCACATNTCC 4122
1760 heAlaGlnValaAspLeuSerTyrThrThrHisLeuGlyGlu...PheSer 1775
4123 ATCANCCTTATTATTTNAGCCTGTCTATACCGATCCGCTTCGGCAAGT 4172
1776 IleThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnG 1791
4173 CCGAACACGCGTCATATCCGCGNGTATGCTCAGGATTTGGCAAAACC 4222
1791 LysIleAsnValSerValTyrAspPheAlaTyrAspValaGlnAsnGln 1808
4223 GCAGTGGGGAATGGCGTAAACCGCGCAATCAAGGTTTCACGCTGTC 4272
1808 InGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1824
4273 NTCCAGCTGCCCGCCGCAAGNCCGCAACTGAGACCGCAACACAGCCG 4322
1825 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1841
4323 GGCATCAATTAAGCTAC 4341
1841 aGluValLysLeuSerPhe 1847

seq_name: p1r2:A81937

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - Neisseria meningitidis
N:Alternate names: Iga proteainase; Iga1 proteinase (EC 3.4.21.7) [misnomer]; Immunog1
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain HF117; strain HF159; strain SM1027
C:date: 05-May-2000 #sequence: revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parikh, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moutre, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; M0ID:20222556
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737
R:Jomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; M0ID:95302961
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 33-548 <LON>
A:Cross-references: EMBL:X82470; NID:g732854; PIDN:CAA57853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown

```

A: Molecule type: DNA
 A: Residues: 53-548 <LOW>
 A: Cross-references: EMBL: X82471; NID: g732858; PIDN: CAA57854.1; PID: g732859
 A: Experimental source: strain HF159
 A: Accession: S61321
 A: Status: nucleic acid sequence not shown
 A: Molecule type: DNA
 A: Residues: 53-548 <LOW>
 A: Cross-references: EMBL: X82472; NID: g732852; PIDN: CAA57855.1; PID: g732853
 A: Experimental source: strain SM1027
 C: Genetics:
 C: Gene: Iga: NMA0905
 C: Superfamily: Iga-specific metalloendopeptidase
 C: Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1410.00 Length: 1849
 Ratio: 1.495 Gaps: 60
 Percent Similarity: 51.001 Percent Identity: 26.501

alignment_block:
 US-09-303-518D-651 x A81937

Align seg 1/1 to: A81937 from: 1 to: 1773

```

49 AAACCGCCGATCCGCTTCGCTGCTTACTTACGATATGCGCTGTC
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2  LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATTTCTCCCAAGCTTGGCGGAGACACTTATTGTCGACATCA 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 atyrAlaLeuThrProTyrSerGluAlaLeuValaIArgAspAspVala 35
149 ACTACCAATACTATGCGACTTGGCGAAATTAAGCAAGTTCAGCTGTC 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 sptYglIlePheArgAspPheAlaGluAsnLysGlyLysPheVal 51
199 GGGCGGAAGATATTGAGGTNTACAAACAAAGGAGGATGGTCGCGCA 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GlYalThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAs 68
249 ATCATGTACAAAGCCGATGATGATTTTCTGTGTCGCGTAC. 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 nAlaLeuSerAsnValProMetIleAspPheSerValaLaAspValaAsn 85
298 ..GGCGTGGCGATTTGGCGGATCAATATATTGATGCGGCGACAT 345
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 rGargThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101
346 AAGCGCGCTATTAACACGTTGATTTGGTCGGAAGAAAGMAATCCGCA 395
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 valLysGlyAspGluIleSerTyrTyrGlnHisAsnGlnLysLeuAs 118
396 T.....CAGCACCGTTTCTTACCAATTTGAAAAAATAATATA 439
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 pValSerAsnAspGluAsnGlnTyrAlaGlnAsnAspTyrG 135
440 AGCCTGACATTC.....CACCTTACACGCG.....GATTAN 474
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 lUpProAsnLysAsnThrHisIleGlnLysGlnGlnIArgLeuGlnAspTyr 151
475 CATATGCGCGGTTTGCATTAATTTGCAGATGACAAACCTGCGCAAT 524
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 AsnMetAlaArgLeuAsnLysPheValThrGluValaIleProIleAlaP 168
525 GACGAGTGACATGAGGGG...AATACCTATTCCGATAAAGAAAAATATC 571
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 oThrSerAlaGlyLysGlnValGlnThrTyrLysAspLysAsnArgPheS 185
572 CCGAGCGGTTCGCGATCGGCTCAGGACACCACTATTGGCTTATGATAT 621
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 erGlnPheValaIArgValaIleGlyAlaIleThrGlnPheGlnTyrAsnSerArg 201

```

```

622 GACAAACACGGCGATTTCCTACTCCGGCCATGTTAATGGCGGCA 671
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 TyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyTh 218
672 TACACATATGCGAG..... 684
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 rProTyrGlnAspValaAsnValThrSerAsnLeuAsnGlnGlnIleuLeu 235
685 ..GGTTGGGAAATTAATGCGCTANTTACTTACGCGGCGATGTCGCCAT 732
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
235 lGlyLysPheGlyAspAsnSerLysHisSerProGlnLysLeuLysGlu 251
733 GCGAACGACTATGGCCCTATGCCATTCGAGTGGCGGCGGACGCG 782
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 ValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAspSerG 268
783 TTCGCAATGTTTATTTATGACAAACAAACATAAATAATGGCTGCTCACG 832
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 ySerProLeuPheAlaTyrAspLysGlnGlnLysArgTyrValPheLeuG 285
833 GAGTTTACAA.....ACGCGTACCTTATTCCGCGAGGAAACGAT 876
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 lYalAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSer 297
877 TTCGAG.....CTGATACGCAAGATTGGTTTACATGATATTTACAG 920
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 TrpGlnGlnTyrPheAsnIleTyrLysGlnPheLysAspIleuLysG 314
921 AGCGATACACATACCGCTTNTTGAACGCGGACGATGACGATTTT 970
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 nArgAsp...AsnAlaGlyThrIleLysGlyTyrGlnHisIleTrpL 330
971 CTTTATATCCACACACACGATGAGGTACGATGATGACAAACCAAGAA 1020
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 yThrThrGlnTyrAsnSerHisIleGlySerThr..... 341
1021 AAGGTNTCCAAATCCAAAGCTTAAAGTACACAGTCCGATGTTTACGA 1070
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 .....AlaValArgLeu..... 345
1071 ATCTTGAATGAACCTGATTAAGAACACGATTACGCGGCGGCGGTATA 1120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 .....AlaGlyAsn... 348
1121 ATCAGTACCGTCCAGGTTAACAACGCTGAACCTTTCTTTATGAT 1170
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe..... 360
1171 TACGGCAACGCAACATCATCTATCAACAAACATCAACCAAGCGCGG 1220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnIleAla 377
1221 CGGTTTGTATTTTGAAGTGATTTTACGGTTCGCGCTGAAACCAAGAA 1269
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
377 yGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyI 394
1270 ..ACGTGGCAAGGCGGCGCTTCATATCAAGTGAAGACGATACCTTACT 1317
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
394 lThrTrpLeuGlnIleAsnGlyIleAspValaIleAspLysLysValaI 410
1318 TGGAAAGTAAACGCGCTGCGCAACGACCGCTGCTCCAAATGCGCAAG 1367
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 TrpGlnValLysAsnProAsnGlnLysAspArgLeuAlaLysIleGlyLysG 427
1368 CACGCTGCACGTTCAAGCCAAAGGGAACCAAGGCTGATCAACGTTGG 1417
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
427 yThrLeuGlnIleAsnGlnTyrThrGlyValaAsnGlnGlnLeuLysValG 444
1418 GCGAGGTACAGTCAATTTGATCAGACGCGGACGAGATTAAGGCAAAAA 1467
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
444 lYAspLysThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLysVal 460

```

```

1468 CAACCTTTAGTGAATCGCTTGNTACGGCGGAGGTACGGTGCACACT 1517
      |||||||..... |||||||..... ||
461 GlnIaIhSerGlnValGlyIleValSerGlyArgGlyThrLeuValLe 477
1518 GAATGCCGATAGTTCACACCCGACAAACCTATTTTCGGCTTCGGC 1567
      |||||..... |||||..... |||||..... |||||.....
477 uAsnSerProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPheArg 494
1568 GCGGACGTTGGATTAAACGGGCATTCGCTTCCTCCACCGCTATTCAA 1617
      |||||||..... |||||||..... |||||||..... |||||||.....
494 IyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHisIleArg 510
1618 AATCCGATGAAGGGGATGATGNCNATCATATGCGCAACACATC 1667
      |||||..... |||||..... |||||..... |||||.....
511 AsnValAspGlnGlyAlaArgIleValAsnHisAsnThrAspArgAlaSe 527
1668 CACCGTTACCATTCACAGGGAATGAAGTATTACAA..... 1704
      |||||..... |||||..... |||||..... |||||.....
527 rThrlleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuS 544
1704 ..... 1704
544 ervaIlyrGluIleArgAsnAspTyrAspAspAspTyrTyrGlyTyr 560
1705 ..... CCGAGTGTAGATATCATAGACTTAA 1733
      |||||..... |||||..... |||||..... |||||.....
561 TyrSerTyrArgLysProIleProGlnGlyLysAspLeuTyrTyrLysAs 577
1734 TTAC..... 1737
      |||||..... |||||..... |||||..... |||||.....
577 nTyArgTyrTyrAlaLeuLysSerGlySerValAsnAlaProMetP 594
1737 ..... 1737
594 roGlnAsnGlyIleThrGlnAsnAsnAspTyrIleLeuMetGlySerThr 610
1738 ..... AGCAAGAAATTCCTTACAC..... 1758
      |||||..... |||||..... |||||..... |||||.....
611 GlnGlnGluAlaLysLysAsnAlaMetAsnHisLysAsnAsnGlnArgI 627
1759 ..... GGTGGTTTGGCGAGAAAGATACGACCAAAAGCAAG 1795
      |||||..... |||||..... |||||..... |||||.....
627 eSerGlyPheSerGlyPhePheGlyGlnGlnAsnGlyLysGlyHisAsn 644
1796 GCGCGCTACACCTGTTTACAGCCGCGCGAGAGACCGACCCGCGT 1845
      |||||..... |||||..... |||||..... |||||.....
644 IyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaGlnAsnArgPheLeu 660
1846 GTTTCGCGGAGCAAAATTTAAAGCGACATCAGCCAAACAGCGCA 1895
      |||||..... |||||..... |||||..... |||||.....
661 LeuThrGlyLysThrAsnLeuAsnGlyLysIleSerValThrGlnGlyAs 677
1896 ACTGTTTTCAGCGGAGCGACAGCCGACCGCTACATCATTTAGGAA 1945
      |||||..... |||||..... |||||..... |||||.....
677 nValLeuLeuSerGlyArgProThrProHisAlaArgAspPheValAsnL 694
1946 GCGGTGCTCAAAATGGAAGT..... ATCCACAGGAGAAATCGTG 1989
      |||||..... |||||..... |||||..... |||||.....
694 ySerSerIaGlnLysAspAlaHisPheSerLysAsnAsnGluValVal 710
1990 TGGGACACGACTGGATCAGCCGACGCTTTAAAGCGGAAATTTCCAT 2039
      |||||..... |||||..... |||||..... |||||.....
711 PheGlnAspAspTrpIleAsnArgThrPheLysAlaThrGluIleThrVa 727
2040 T... CAGGGCGGAGCGGCGGTGATTTCCGCAATGTTGCCAAGTGAAG 2086
      |||||..... |||||..... |||||..... |||||.....
727 IAsnGlnSerAlaSerPheSerSerGlyArgAsnValSerAsnIleThrA 744
2087 GCGATTGNCATTTGAGCAATCAGCCGACGATTTTGTGTCGACCG 2136
      |||||..... |||||..... |||||..... |||||.....
744 IAsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr..... 758
2137 CATCAAGCCATACATCTGTACAGCTTCGAGCTGACNGCTGTACAAA 2186

```

```

759 LysAsnGlyAspGluValCysValArgSerAspTyrThrGlyTyrValTh 775
      |||||..... |||||..... |||||..... |||||.....
2187 TTGTGTGCANAAANCAATTACCGACGATTAAGTGTGCTTCATTCAGTA 2236
      |||||..... |||||..... |||||..... |||||.....
775 rCyAsnThrGlyAsnLeuSer... AspLysAlaLeuAsnSerPheGlyA 791
2237 AGCAGNCATNTNAGCGGANTGTNAGCTNCCNATNAGCTNTNTNAAAN 2286
      |||||..... |||||..... |||||..... |||||.....
791 IaThrLysIleAsnGlnLysAsnValAsnLeuAsnGlnAlaAlaLeuVal 807
2287 CTCNCGGCGCTGCTCANNACTNANAGCAATCTTACTGCAATGCGATAC 2336
      |||||..... |||||..... |||||..... |||||.....
808 Leu... GlyAlaAlaLeuThrGlyGlnIleGlnGlyGlnLysAsnSe 823
2337 ACCTTTACAGTCAGCCACACAGCC..... ACCC 2365
      |||||..... |||||..... |||||..... |||||.....
823 rArgValSerLeuAsnGlnHisSerLysTrpHisLeuThrGlyAspSerG 840
2366 AAACCGCAACCTTACCTC..... CTG 2388
      |||||..... |||||..... |||||..... |||||.....
840 IValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAla 856
2389 GGCATATGCCACAGCAACATTTAATCAAGCCACATTTAAAGCGACNCAATC 2438
      |||||..... |||||..... |||||..... |||||.....
857 SerAspAlaGlnSerAlaAsnGlnTyrHisThrLeuLysIleAsn... H 872
2439 GGNITCGGCAATGCTTCATTTAATCTAAGCAACAGCCGACCAAMC 2487
      |||||..... |||||..... |||||..... |||||.....
872 sLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAlaGluAsnL 889
2487 ..... 2487
889 euGlyAspLysValLeuValLysGluSerAlaSerGlyHisTyrGlnLeu 905
2488 ..... GCGAGTGTGAGCGCTTC 2504
      |||||..... |||||..... |||||..... |||||.....
906 HisValGlnAspLysThrGlyLysProAsnGlnGlyLeuAsnLeuH 922
2505 CGACACAGCT..... AAGCAAAAGTAAAGCATTCGCGACTCA 2542
      |||||..... |||||..... |||||..... |||||.....
922 eAspAlaSerSerValArgAspArgSerHisLeuSerValSerLeuAla 939
2543 ACGGCAATGTCTCCCTTAAGCCGATTAAGCAAGTATTCATTT..... 2581
      |||||..... |||||..... |||||..... |||||.....
939 snHisHisValAspLeuGlyAlaLeuArgTyrThrIleLysThrGlnAs 955
2582 ..... TTGA 2585
      |||||..... |||||..... |||||..... |||||.....
955 nGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnArgArgValL 972
2586 AAACAGCGCTTACCGGACACT..... 2609
      |||||..... |||||..... |||||..... |||||.....
972 ySProAlaProSerProAlaThrAsnThrAlaSerGlnAlaGlnThrAs 988
2609 ..... 2609
988 pSerAlaGlnIleAlaLysProGlnAsnIleValValAlaProPheSer 1005
2610 ..... CAGCGGAGCAAGCAAGACAGCATT 2633
      |||||..... |||||..... |||||..... |||||.....
1005 roGlnAlaAsnGlnAlaGlnGluAlaLysArgGlnGlnAlaLysAlaGlu 1021
2634 ACACCTTAAAGACAGCAATGAC..... 2657
      |||||..... |||||..... |||||..... |||||.....
1022 GlnValLysArgGlnGlnAlaGlnAlaGlnAlaGlnArgLysSerAlaGluLeuAl 1038
2658 ..... CTTGCCGTCAGGACAGCAATTTAGGCAATTTAAACCTTGAC 2697
      |||||..... |||||..... |||||..... |||||.....
1038 aLysGlnLysAlaGlnAlaGlnArgGlnAlaArgGlnLeuAla..... T 1053
2698 AACGCCACCATTTACATTCATTCGCTTATCGCCAGCATCTCGAGCGC 2747
      |||||..... |||||..... |||||..... |||||.....

```

1053 hrArgGlnLysAlaGlnGlnGlnLysSerSerAlaGlnLysAlaArgArg 1069
2748 GCAACGGCGGNGTGTACAGACAGCGCGCGCGCGTTCGCGCGTCC 2797
1070 HisGlnLysGln.....ArgGlnAlaAlaGlnLysSerAlaLysGln 1084
2798 TATT.....ATCGGTACACGGCA 2817
1084 sValGlnAlaGlnLysArgGlnAlaGlnAlaValAlaArgLysAlaG 1101
2818 ACTTCGGTACATCCGTTTCACACCGCTGACGGTAAACGGCAATTGAA 2867
1101 LysAlaGlnGlnLysArgGlnAlaAlaGlnLysAlaArgGlnHisGln 1117
2868 CNGTCAGAGAACATTCGCTTTATGTGGAACTTCGGCTA..... 2909
1118 LysGlnLysArgGlnAlaAlaGlnLysSerAlaLysGlnArgValGlnGln 1134
2910 ..CCGACGCAACAATTGCAAGCTGCGGAAAGTTCCGAAAGNACTTACAC 2958
1134 uGlnArgArgGlnThrAlaGlnSerGlnProGlnArgArgLysArgArg 1151
2959 TTGGCGGTCAACAATAC.....CGGCAAGCAACCGGT 2990
1151 LaAlaProGlnAspTyrMetAlaAlaSerGlnAspArgProLysArgArg 1167
2991 AAGCTTCATCAATTGACGGTAGTGGAAAGGAA..... 3023
1168 GlnHisArgSerValGlnGlnAsnAsnValGlnLysAlaGlnAlaGln 1184
3024AGACACAA..... 3032
1184 aGlnLysAlaArgArgGlnGlnGlnGlnLysAlaAlaGlnLysGln 1201
3033ACCGTGTTC 3041
1201 LaLysGlnArgAlaGlnLysArgGlnAlaGlnAlaLysAlaAlaArg 1217
3042 CGAAACCTTAATTTCACCGTGCACAAACGACAGCTGATGCCGCGCGCT 3091
1218 ArgLysAlaGlnLysArgGlnLysAlaArg..GlnAlaAlaGlnLysAlaGln 1234
3092 GGGTTACCAACTCATCCGCAAGAGCGGAGTTCGCGCTGCATTAAT... 3138
1234 IsArgGlnGlnLysAlaGlnLysAlaAlaGlnLysSerAlaAsnGlnLys 1250
3139 ...CCGTCACAAAGACAGAGCTTCGACAAACTCGCAAGCGCAAGAGC 3185
1251 AlaAlaAlaGlnLysAlaGlnAlaLysAlaAlaArgGlnGlnLysAlaLys 1267
3186 CAAAAACAGCGCGGAAAGACACAGCGCAAGCGCTT..... 3222
1267 aArgGlnGlnGlnLysAlaArgLysAlaAlaGlnLysAlaValLysGln 1284
3223GACGGCTGATTCGCGCGCGCGCGCGCATGCC 3252
1284 yAlaGlnLysArgLysThrAlaGlnLysAlaLysGlnArgAlaAla 1300
3253 GCCGAA.....AAGACGAAAGCGTTGCCGCAACCGC 3284
1301 AlaGlnAlaAlaLysArgGlnGlnLysAlaArgGlnThrAlaGlnLys 1317
3285 CCGGNGCGAGCGCGGAAATGTGGCATTTGACAGCGGAGAA..... 3330
1317 aArgArgGlnGlnLysAlaArg.....GlnAlaAlaGlnLys 1330
3331 ..GACAAAAACGGGTGCAGCGGATTAAGACAGCGGNTGCGGAAACAG 3378
1330 eAlaLysGlnLysAlaGlnLysArgGlnLysAlaAlaGlnLysAlaLys 1346
3379 CGCGAAGCGGAAACCGCGCGGNTACACCGCTTC..... 3414
1347 ArgLysAlaGlnGlnLysAlaArgGlnLysAlaAlaGlnLysGlnProGln 1363
3414 3414
1363 naArgLysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGln 1380
3415CCCGCGCGCGCGCGCGCGCGCGGATTTGCGGCAACG... 3453
1380 snArgProLysArgArgGlnLysArgSerThrLeuProAlaProProser 1396
3453 3453
1397 ProSerPheAspSerSerAlaTyrAlaAlaProAlaAlaLeuHisAsnPr 1413
3453 3453
1413 oAspTyrGlnLysAsnAspTyrGlnLysLeuLeuProLeuAspAlaLeuGln 1430
3453 3453
1430 sPGLysAsnValSerGlnLysValAspThrSerAspLysGlnProGlnAsp 1446
3453 3453
1447 AsnThrGlnLysHisGlnLysValGlnLysValSerLeuGlnProAlaGln 1463
3454 ...CAGCC...CAACCGCAACCTCAACCGCAAC 3481
1463 aAlaGlnProAlaGlnAlaAlaThrGlnProGlnAlaGlnAlaAla 1480
3482 GCGACGGGACCGTGTATNAGCGGTATGCGCATACCGGTTGAGCAATT 3511
1480 LaGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLysSerAla 1496
3532 TCAGCGACGTCACACAGCTTTTCGCC...GTACAGACGAATTGACAGC 3578
1497 MetAlaSerThrGlnSerThrLeuLeuAspThrGlnLysAlaSerLeuThr 1513
3579 CGGTTCGCGCAAGACCGCCG...AACGCGTTTGA 3613
1513 GlnHisLeuLysSerArgAlaAspAlaGlnLysAsnSerValLysPro 1530
3614 CAAGCGCATCCGCGNACACAAACATACAGCTTCGCAAGATTCGCGGCC 3663
1530 eSerAsnLysGlnLysArgAspTyrAlaSerAlaGlnLysArgArg 1546
3664 TACCGCCACAAACCGACCTGCGCAATCGGTATGCAAAAACTCGG 3713
1547 PheSerSerLysArgThrGlnThrGlnLysLeuAspArgSerLeu 1563
3714 CAGCGGG...CGGTCGCGCATCCGTTTTCGCAACACCGGACGCAACA 3760
1563 GlnAsnMetGlnLysGlnLysAlaLeuThrTyrSerAspSerLeuHis 1580
3761 NCTTCAGACGAGCGATCGCAACTCGGACAGCGCTTCGCGCGCGCTT 3810
1580 hrPheAspGlnAlaSerGlnLysAsnThr...PheValGlnAlaAsnLeu 1595
3811 TTGCGGAATACGCGATCGGACGTTGCACATCGGCATCGACAGCGGCGC 3860
1596 TyrGlnLysTyrTyrLeu.....AsnAspAl 1604
3861 GGGTTTACGAGCGCGCATCTTCAGACGCGCATCGGAGCAAAATCGCG 3910
1604 aTyrTyrMetAlaGlnLysPheLeu.....AlaGlnSerLeuArg 1618
3911 GCGCGGTCTGCATTTACGCGATTCAGGACAGATACCGCGCGGTTTC... 3957
1618 eArg.....LeuGlnThrGlnGlnLysAlaAsnPheAsn 1629
3958CG 3959
1630 ArgThrSerThrGlnThrGlnLysLeuGlnLysAlaAsnThrLeuLys 1646


```

1444 SerAspLysGlnProGlnAspAsnThrGlnLeuHisGluLysTyrGluAs 1460
3462 ..... 3462
1460 msAspTyrGluGluIleProLeuAspAlaLeuGlnAspGluAspValSerg 1477
3463 .....CCGCAACCTCA..... 3474
1477 LuSerValAspThrSerAspLysGlnProGlnAspAsnThrGlnLeuHis 1493
3475 .....CCCAACCG..... 3483
1494 GluLysValGluThrValSerLeuGlnProAlaAlaGlnProArgAl 1510
3484 .....CAGCGCGACCTGA 3496
1510 aglnAlaIaThrGlnLeuGlnAlaGlnAlaAlaIaGlnAlaAspAlaVal 1527
3497 TNAGCCGTTAGCAATAGCGGTTTGAATTTCCGCGACCGCTCAAC 3546
1527 alSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGln 1543
3547 AGCGTTTCCGC...GTACAGACCAATTGACCGCGTGTTCGCGAAGA 3593
1544 SerIleLeuLeuAspThrGlyAlaSerLeuThrArgHisIleAlaGlnLly 1560
3594 CGCGCCG.....AACGNGTTTGACACAGCGACGACATCCGCGN 3628
1560 sSerThrAlaAspAlaGluLysAsnSerValAlaTrpMetSerAsnThrGlyT 1577
3629 ACACCAAACTACCGCTTCGCAAGATTCCGCGCTACCGCCCAACCAACC 3678
1577 YRGLYArgAspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysArg 1593
3679 GACCGCGCCAAATCGGTATGACAAAACTCGCGACGCGG...CGCGT 3725
1594 ThrGlnThrGlnIleGlyIleAspArgSerLeuSerGluAsnMetGlnI 1610
3726 CGGATCCGTTTTCGACAAACCGGACGCAAAACNTTCGACGACGCA 3775
1610 eGlyGlyValLeuThrTyrSerAspSerGlnHisThrPheAspGlnAlaAs 1627
3776 TCGGCACTCGGACGCGCTTCGCCACGCGCGCTTTTCGGCAATACGCG 3825
1627 eGlyLysAsnThr...PheValGlnAlaAsnLeuTyrGlyLysTyr 1642
3826 ATCGGC...AGTTTCGACATCGGCATCAGCACGGCGCGGTTTTCAGCAG 3872
1643 LeuAsnAspAlaIaTrpTyrValAlaGlyAspIleGlyAlaGlySerLeuAr 1659
3873 CGGCACTTCATTCAGACGCGCATCGGCAAAATCCGCGCGCGGCGCGC 3922
1659 gSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgAlaSerIleG 1676
3923 ATTACGGCATTCAGGACGATACCGCGCGGTTTCGGCGGATTCGGCATC 3972
1676 InThrGlyLeuThrLeuGlnLysAsnThrLeuLysIleAsnGlnPheGlnIle 1692
3973 GACCGTCATTCGGCGCAACCGCTATTTCGCCAAAAAGGGATTAC 4022
1693 ValProSerAlaGlyIleArgTyrSerArgLeuSerAlaAspTyrLys 1709
4023 CTACGAAACGTCATATCGCCACCGCGGTCGTTCAACGCGNATAC 4072
1709 sLeuGlyLysAsnAspSerValLysValSerSerMetSerValLysThrLeuT 1726
4073 GNGCGGGATTAAAGCAGATTATTCATTAACCGCGCAACAGATATTC 4122
1726 hrAlaGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuThr 1741
4123 ATCAGNCCTTATTNAGCCTGTCTAT...ACCATGCGCGCTTCGGCGAA 4169

```

```

1742 ValLysProLeuLeuSerAlaAlaTyrPheAlaAsnTyrGlyLysGly 1758
4170 AGTCGGAACAGCGCTCAATACCGCGNATTTGGCTCAGGATTTGGCAAAA 4219
1758 yValAsnValGlyLysAsnSerPheValTyrLysAlaAspAsnGlnGln 1775
4220 CCCGCACTCGCGGAATGGGCGCTAAACCGCGAATTCAAAGGTTCCAGCG 4269
1775 InTyrSerAla.....GlyAlaAlaLeuLeuTyrArgAsnValThrLeu 1789
4270 TCCNTCCAGCGTCGCCCGCGCAAGCGCAACCTGACGACGCAACAG 4319
1790 AsnValAsnGlySerIleThrLysGlyLysGlnLeuGlnLysGlnLysSe 1806
4320 CGCGGCGATCAAAATTAGCTACCGCTCG 4347
1806 rGlyGlnIleLysIleGlnIleArgPhe 1815

```

seq_name: p1r2:154632

seq_documentation_block:
 tsh protein - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Dec-2000
 C:Accession: 154632
 R:Providence, D.L.: Curtiss, R.
 Infect. Immun. 62, 1369-1380, 1994
 A:Title: Isolation and characterization of a gene involved in hemagglutination by an
 A:Reference number: 154632; M01D:94178945
 A:Accession: 154632
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1377 <RFS>
 A:Cross-references: GB:127423; NID:9469235; PIDN:AAA24698.1; PID:9469236
 C:Superfamily: Iga-specific metalloendopeptidase

alignment_scores:
 Quality: 1084.50 Length: 1508
 Ratio: 1.378 Gaps: 54
 Percent Similarity: 52.188 Percent Identity: 25.995

alignment_block:
 US-09-303-518D-651 x 154632 ..

Align seg 1/1 to: 154632 from: 1 to: 1377

```

90 ATGCGTGTCTGTCGCAATTCCTCCCAAGCTTGCGCGGACACACTTATT 139
   :|||||: 111111: |||||:|||||: 33
42 IleProValLeu...PheSerAlaGlySerLeuAlaGlyThrValAsnA 57
140 TCGGCATCAACTACCAATACATACGCGACTTTCGCCAAATAAAGCAAG 189
   :|||||:|||||: |||||:|||||:|||||:
57 snGluLeuGlyTyrGlnLeuPheArgAspPheAlaGlnLysnLysGlyMet 73
190 TTTCAGTTCGGGCGAAGATATTGAGTWTACAAACAAAAGGGAGATT 239
   ||| ||||| :|||: |||||:|||||:|||||:
74 PheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGlnIleGlu 90
240 GGTGCGCAAAATCAATGACAAAGCCCGGATGATTGTTTCTGTGTGTG 288
   :|||||: |||||: |||||: |||||: |||||:
90 eValGly...ThrLeuAspLysAlaIaMetProAspPheSerAlaValA 106
289 ..TCGCGTAACGCGGTGGCGCATTTGGTGCGCATATATATTGTGAC 336
   |||||: |||||: |||||: |||||: |||||:
106 sPserGluIleGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSer 122
337 GTGCGCATTAACGGCGGTATACACAGTTGATTTGGCGGAAGAGAG 386
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 ValLysHisAsnGlyLysTyrThrAsnValSerPheGly...AspGly 138
387 NAATCCGATCAACACCGTTTTCATCAAAATTGTAAGAATAAT 436
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

138 uasn.....ArgTyrAsnIleValAspArgAsnAsnA 149
437 ATTAGCGTCGACAAATTCACACCTTACAAAGCGGATTANCATATGCGCGGT 486
149 laPro.....SerLeuAspPheHisAlaProArg 158
487 TTGCATTAATTTGTGCAGATGCAGAAACCTGTGGAATGACGAGTAC.. 534
159 LeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaGlnG 175
535 .ATGAGGGGAATFACCTATTCGATTAAGAAAAATATCCCGAGCGTGC 583
175 ValValAlaGlyAlaTyrLeuAspLysGluArgTyrProValPheTyr 192
584 GCATCGCGTCAGACACACACTATGCGCTATGATGATGACAAACAGCG 633
192 rGlnGlySerGlyThrGlnTyrIleLys.....AspSerAsnGly 205
634 GATTATCTCTACTCCGCG.....GCATGTTAATTGGCGGCAATAC 674
206 GlnLeuThrGlnMetGlyGlyAlaTyrSerTrpLeuThrGlyGlyTyrVa 222
675 ACATATGCGAGGTGGGGAATAATAGC.....GTANTACTTTGACGG 718
222 LglSerLeuSerSerTyrGlnAsnGlyGlnMetIleSerTrpSerSerg 239
719 GCGATGTGCGCCATGCCAAGCAACTATGCGCTATGCGCATGTCAGTGG 768
239 LysLeuAlaPheAspTyrLysLeuAsnGlyAlaMetProIleTyrGlyG 255
769 GCAGGCGACAGCGGTTCCGCAATGTTTATTATGACAAACAAACAATRA 818
256 AlaGlyAspSerGlySerProLeuPheAlaPheAspTrpValGlnAsn 272
819 ATGCGTGCACAGCGAGTTTACAAACGCGTACCCTTATCCGCGAGG 868
272 strPValIleValGlyValIleuThrAlaGlyAsnGlyAlaGlyLys 289
869 AAAACGGTTTCCAGCTGATACGCMAAGATGGTTCTACGATGACATTAC 918
289 LysAsnTrpAlaValIleProLeuAspPheIleGlyLysPheAsn 305
919 AGAGGGGATACACATACCGTCTNTTTTGAACCGCGACGTAACGACATT 968
306 GluAspAsnAspAlaProValThrPheArgTrpSerGlnGlyAlaLe 322
969 TTCCCTTACATCCAAACAAC...GGTACGGGTACGTAACAAACA 1015
322 uGluTrpSerPheAsnSerSerThrGlyAlaGlyAlaLeuThrGlnGly 339
1016 ACGAAAAAGTNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTT 1065
339 hr..... 339
1066 GAGCATCTTTGATGAACACTGATAAGAACCAGTTTACGGCGAGGGG 1115
340 .....ThrThrTyrAlaMetHisG 346
1116 TGTTAATCAGTACCGTCCAAAGTTAAACAACGCTGAAAACTTTCTTTA 1165
346 Y.....GlnGlnGlyAsnAspLeuAsnAlaGlyLysAsnLeuIlePhe. 360
1166 TCGATTACGGCAACGCAAACTCATCTTATCAACAACATCAACCAAGCG 1215
361 ..GlnGlyGlnAsnGlyGlnIleAsnLeuLysSerValSerGlnGly 376
1216 GCGCGGGGGTTGTATTTGATGAGTGAATTTAGCGTCTGCCGGAACAA 1265
377 AlaGlySerLeuThrPheArgAspAsnTyrThrValThrThrSerAsn 393
1266 CGAAGCTGCGCAAGCGCGGCTTCATCATGACGTAAGACAGTACCGTTA 1315
393 ySerThrTrpThrGlyAlaGlyIleValValAspAsnGlyValSerVal 410
1316 CTTGAAAGCTAAACGGCGTCGCAACGACGCGCTGTCCAAATCGCCAA 1365
410 snTrpGlnValAsnGlyValLysGlyLysAsnLeuHisLysIleGlyG 426
1366 GGCACGCTGCACGTTCAAGCCAAAGGGAAACCAAGCGCTGCATCAGCT 1415
427 GlnThrLeuThrValGlnGlyThrGlyIleAsnGlnGlyGlyLeuLys 443
1416 GCGCGACGCTGACAGTCAATTTGATCGACGACGACGATTAAGGCAAA 1465
443 LglYAspGlyLysValValLeuAsnGlnGlnAlaAspAsnLysGlyG 460
1466 AACAGCCTTTAGTGAATCGCCTTGTCAGCGGACGCGTACGCTGCAA 1515
460 alGlnAlaPheSerSerValAsnIleAlaSerGlyArgProThrVal 476
1516 CTGAATCGCATATACGTCAACCCCGACAAACTCATTTCCGGCTTTC 1565
477 LeuThrAspGluArgGlnValAsnProAspTrpValSerTrpLysTyr 493
1566 CGCGGACGCTTGTGATTTAAACGGGACATGCGTTCCTTCACCGTATTC 1615
493 gGlyGlyThrLeuAspValAsnGlyAsnSerLeuThrPheHisGlnLeu 510
1616 AAAATACCGATGAAGGGCGCATGATTCNATCATTAATGCCACACACAA 1665
510 yAlaAlaAspTyrGlyAlaValLeu...AlaAsnAsnValAspLysArg 525
1666 TCCACCGCTTACCATTCAGGGAATGAAGTATTACACAAACGAGTGTA 1715
526 AlaThrIleThr..... 529
1716 GAATATCATATGACTTAATTAC.....ACCAAGAAATTCCTACA 1756
530 .....LeuAspTyrAlaLeuArgAlaAspLysValAlaLeu 542
1757 ACGGTTGG..... 1764
542 snGlyTrpSerGlnSerGlyLysGlyThrAlaGlyAsnLeuTyrLysTyr 558
1765 .....TTGGCGGAAAGATACGAC 1784
559 AsnAsnProTyrTrpAsnThrThrAspTyrPheIleLeuLysGlnSer 575
1785 C..... 1785
575 rTyrGlyTyrPheProThrAspGlnSerSerAsnAlaThrTrpGluPhe 592
1785 ..... 1785
592 alGlnHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsn 608
1786 .....AAAACGACGGCGGCTCAACCTGT 1811
609 ThrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnVal 625
1812 TTACCAG...CCGCGCGCAGAGACCGCACCCGCTCTTCGCGGGA 1858
625 pAsnArgLeuProGlnGlyValThrGlyAlaLeuValMetAspLysAla 642
1859 CAATTTTAACGGCAACATCCGCAACAAACGCAACGTAAGCTTTTCAGC 1908
642 laAspIleSerGlyThrPheThrGlnGlnAsnGlyArgLeuThrLeuGln 658
1909 GCGAGACGCAACGCGCTTACATCAT..... 1938
659 GlyHisProValIleHisAlaTyrAsnThrGlnSerValAlaAspLys 675
1939 .TTAGGAACGCGGTGTCAAAATGGAAGTATCCCAAGAGGAAATCG 1987
675 uAlaAlaSerGlyAspHisSerVal.....LeuThrGlnProThrSer 690

```

```

1988 TGTGGCAACAGCACTGGATCMACCGACGTTTAAAGCGGAATTTTCAT 2037
      ::::::::::::::::::::
690 heserIngluAspTrpGluAsnArgSerPheThrPheAspArgLeuSer 706
      ::::::::::::::::::::
2038 ATTCAAGGCGGCGACGGCGGATTTCCCGCAATGTT.....GCCAA 2078
      ::::::::::::::::::::
707 LeuIlyAsnThrAspPheGlyLeuGlyArgAsnAlaThrLeuAsnThr 723
      ::::::::::::::::::::
2079 AGTGAAGGCGATTCGCAATTTGAGCAATGCACCGCCAGCGATTTTGTG 2128
      ::::::::::::::::::::
723 rIleGlnAlaAsp.....AsnSerValThrLeuGly 734
      ::::::::::::::::::::
2129 TCCGACCCGATCAACGCCATACATGTGACAGCTGGAGCTGAAGT 2178
      ::::::::::::::::::::
735 .....AspSerArgValPheIleAspIlyAsnAspGlyGlnGly 747
      ::::::::::::::::::::
2179 CTGCAAAATGTGTGCA.....NAANCAATTACCGACGATTAAGT 2219
      ::::::::::::::::::::
748 ThrAlaPheThrLeuGluGluGlyThrSerValAlaThrIlyAsp.... 762
      ::::::::::::::::::::
2220 GATTCCTTCATGTACTAGACAGCAGTACCGGAGCTGAGCTGAGCTG 2269
      ::::::::::::::::::::
763 .....AlaAspIlySerValPheAsnGlyThrValAsnLeuAsp 776
      ::::::::::::::::::::
2270 ATTAGCANTNTTNAANCTCNCNGGCGTGCNNCACTNAANGCAATCTT 2319
      ::::::::::::::::::::
776 snGlnSerValLeuAsnIleAsnAsp.....IlePheAsnGlyIlyle 790
      ::::::::::::::::::::
2320 AGTGCAAATGGCGATACAGCTATACAGTACGCCACACAGCCACCCAAA 2369
      ::::::::::::::::::::
791 GlnAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804
      ::::::::::::::::::::
2370 CGGCAACCTTAGCCTGTGGGCAATGCCACAGCAATTTAATCAAGCCA 2419
      ::::::::::::::::::::
805 .....ValleuGlyAsnSerThrLeuThrSerThrAlaLeu 817
      ::::::::::::::::::::
2420 CATTTAAAGCGCAACNCATCGGNTTGGCGCAATCTTCAATT..... 2460
      ::::::::::::::::::::
817 snLeuAsnIlyGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAsp 833
      ::::::::::::::::::::
2461 .....AATCTAAGCAACACAGCCGCACAA..... 2484
      ::::::::::::::::::::
834 GlyProValAsnIleSerAspAlaAlaLeuSerLeuAsnSerArgProAs 850
      ::::::::::::::::::::
2485 .....AACGCAAGTCTGA 2497
      ::::::::::::::::::::
850 peIuValSerHisThrLeuProValTyrAspTyrAlaGlySerTrp 867
      ::::::::::::::::::::
2498 CGCTT...TCGACAGACGCTTAAGCAACGTAAGC...CATTCGCACTC 2541
      ::::::::::::::::::::
867 snLeuIlyGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeu 883
      ::::::::::::::::::::
2542 AACGGCAATGTCTCCCTAGCCGATAG..... 2568
      ::::::::::::::::::::
884 SerGlyAsnIleAsnValGlnAspIlySerValThrValThrLeuGlyGly 900
      ::::::::::::::::::::
2569 .....GCAG 2572
      ::::::::::::::::::::
900 uGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSerL 917
      ::::::::::::::::::::
2573 TATTCATTTGAAAACAGCCGCTTTACCGGACAACGAGCGGACGACAG 2622
      ::::::::::::::::::::
917 eupheAsnGlyTyrArgAsnIleTrpSerGlySerLeuAsnAlaProAsp 933
      ::::::::::::::::::::
2623 GANACAGCATTTACACTTAAAGACAGCAATGACGCTGCGTGACGAC 2672
      ::::::::::::::::::::
934 AlaThr...ValSerMetThrAspThrGlnTrpSerMetAsnGlyAsn 949
      ::::::::::::::::::::
2673 GCAATTTAGCAATTTAAACCTTGACAAAGCCACCATTAACACTCAATCCG 2722
      ::::::::::::::::::::
949 rThrAlaGlyAsnMetIlyLeuAsnArgThrIleValAlaGlyPheAsnIly 966
      ::::::::::::::::::::
2723 CCAATCCGCAAGATGCTGACAGCCGCGCAACCGGACAGNGTGTGACAGC 2772
      ::::::::::::::::::::

```

```

966 Iy..... 966
      ::::::::::::::::::::
2773 CCGCGCCGCGCTTCGCGCGCTTCCTATTATCCGTTACACCGCAACTTC 2822
      ::::::::::::::::::::
966 ..... 966
      ::::::::::::::::::::
2823 GSTNAATCCCGTTTCAACAGCTGACGTTAAGCGCAATTAAGACGTC 2872
      ::::::::::::::::::::
967 .....TherProPheThrThrLeuThrThrAspAsnLeuAspAlaValG 982
      ::::::::::::::::::::
2873 AAGAAATTCGCGCTTATGTGCAAGTCTGCGCTACCGGAGGACAAA 2922
      ::::::::::::::::::::
982 InSerAlaPheValMetArgThrAspLeu.....AsnIlyAlaAspIly 996
      ::::::::::::::::::::
2923 TTGAAGCTGGCGGAAAGTTCCGAAGGAACTTACACTTGGCGGTCAACA 2972
      ::::::::::::::::::::
997 LeuValIleAsnIlySerAlaThrGlyHisAspAsnSerIleTrpValAs 1013
      ::::::::::::::::::::
2973 TACCGGCAACGACCCGTAACCTCGATCAATTG..... 3006
      ::::::::::::::::::::
1013 PheLeuIlySerProSerAsnIlyAspThrLeuAspIleProLeuValS 1030
      ::::::::::::::::::::
3007 .....ACGTA 3012
      ::::::::::::::::::::
1030 erAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgVal 1046
      ::::::::::::::::::::
3013 GTGGAAGGGAAGAACAAACCGCTGTCCGMAAACCCTTAATTTCAACCT 3062
      ::::::::::::::::::::
1047 ValGlyPheSerAspValThrProIleLeuSerValAlaArgIysGluAspG 1063
      ::::::::::::::::::::
3063 GCAAAAGCAACAGTGCATGCGCGCGGTGCGCTTACCAACTATCCGCA 3112
      ::::::::::::::::::::
1063 IlyIysGlyGluTrpValIleuAspGly.....TyrGlnValAlaArg 1077
      ::::::::::::::::::::
3113 AAGAGCGCGATTCGCGCTCATATCCGGTCAAGAAAGAGCTTTC 3162
      ::::::::::::::::::::
1077 snAspGlyGln..... 1080
      ::::::::::::::::::::
3163 GACAAACTCGGCAAGCAGAAACCCAAACAGCGGAAAGAACAC.. 3210
      ::::::::::::::::::::
1081 .....GlyIysAlaAlaIleThrPheMetHisIleSerTyrAsnAs 1094
      ::::::::::::::::::::
3211 .....GCGCAAGCCTTGACGCGCGGCTGATTCGCGCGCGCGCGC 3247
      ::::::::::::::::::::
1094 nPheIleThrGluValAsnAsnLeuAsnIlyAspMetGlyAspLeuArg 1111
      ::::::::::::::::::::
3248 ATGCCCGCGAAAGACAGAAAGCGTTGCCGAAACCGCGCGCGGACG 3297
      ::::::::::::::::::::
1111 spIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnGlySerGly 1127
      ::::::::::::::::::::
3298 GGGGAAATGTCCG.....ATTATGACGCGGAGGA 3329
      ::::::::::::::::::::
1128 SerAlaAspIlyGlyThrPheThrAspHisTyrThrLeuLeuGlnMetGlyAl 1144
      ::::::::::::::::::::
3330 AGAGAAAAAAGCGGTGCAGCGGATTAAGAC..... 3360
      ::::::::::::::::::::
1144 aaSPArgIlyHisGluLeuGlySerMetAspLeuPheThrGlyValMet 1161
      ::::::::::::::::::::
3361 .....AAGCANTTGGCGAAACAGCGGCAAGCGGAAAC 3393
      ::::::::::::::::::::
1161 laThrTyrThrAspThrAspAlaSerAlaAspLeuTyrSerGlyIlyThr 1177
      ::::::::::::::::::::
3394 CGGCGGNTACACACGCGCTTCCCGCGCGCGGNGCGCGCGGATTT 3443
      ::::::::::::::::::::
1178 IysSerTrpGlyGlyIlyPheTyrAlaSerGlyLeuPheArgSerGlyAl 1194
      ::::::::::::::::::::
3444 GCGCAACGCGACGCCAAACCGCAACCTCAACCCCAACCGCGCGAC 3493
      ::::::::::::::::::::
1194 aTyrPheAspValIleAlaIlyTyrIleHisAsnGlnIlyAsnIlyTyrAspL 1211
      ::::::::::::::::::::
3494 TGATNAGCCGTTATGCCAATAGCGGTTTGAATTTCCCGCACCGCTC 3543
      ::::::::::::::::::::

```

```

1211 eu.....AsnPhleAlaGlyAlaGlyLysGlnAsnPhleArgSer..... 1223
3544 AACAGCGTTTTCGGCTACAGAGAA.....TTGGA 3575
      ::::::::::::::
1224 HisSerLeuTyrAlaGlyAlaGlyLysGlnAsnPhleArgSerHisLeuThrAs 1240
3576 CCGCGGTGTTGGCCAGAACCGCCGCAACGCGNTTGGACAAAGNCATCC 3625
      ::::::::::::::
1240 pThrThrPheValGluProGlnAlaGluLeuValTyrGlyArgLeuGln 1257
3626 GGNACCAACCACTACCGGTTCGCAAGATTTCCGCGCTACCGCAACAA 3675
      ::::::::::::::
1257 LysGlnThrPheAsnTyrPheAsnSerGlyMet..... 1267
3676 ACCGACCTGCGCAATCGGTATGCAAGAAAACTCGGCAAC..... 3717
      ::::::::::::::
1268 .....AspValSerMetArgArgAsnSerValAlaAsnProLeuVal 1280
3718 .GGCGCGCTGGCAGTCCTGTTTTCGCAACCGGACGCAAAACANCTTCG 3766
      ::::::::::::::
1280 LysArgThrGlyValVal.....SerGlyLysThrPheSerGlyLysA 1295
3767 ACAGCGCGCATCGCAACTCGGCAAGCTTTCGCGCGCGCTTTCGCG 3816
      ::::::::::::::
1295 sPTPserLeuThrAlaArgAlaGlyLeuHisTyr..... 1306
3817 CAATACGCGCATCGGAGTTCGACATCGCATCGACAGCGCGCGGCTT 3866
      ::::::::::::::
1307 .....GluPheAspLeuThrAspSerAlaAspValHisLe 1318
3867 TAGCGCGCGCATCTTCAGACGCGCATCGCAAAATCCGCGCGCG 3916
      ::::::::::::::
1318 ulysAspAlaAlaGlyLysGlnHisGlnLeuAsnGlyArgLysAspSerArgm 1335
3917 TGCTGTCAT...TAGCGCATTCAGCGCAGCATAC..... 3945
      ::::::::::::::
1335 eLeuTyrGlyValGlyLeuAsnAlaArgPheGlyAspAsnThrArgLeu 1351
3946 .....CGCGCGCGTTCGCGGATTCGCGCATCGCAACGCTA 3980
      ::::::::::::::
1352 GlyLeuGluValAlaGluArgSerAlaPheGlyLysTyrAsnThrAspAspAl 1368
3981 CATCGCGCGCAACG...CGCTAT 3999
      ::::::::::::::
1368 allAsnAlaAsnLeuArgTyr 1375
seq_name: p12:S57664
seq_documentation_block:
IGA-specific metalloendopeptidase homolog sepa precursor - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 19-Oct-1995 #sequence.revision 03-Nov-1995 #text_change 08-Dec-2000
C:Accession: S57664; S69769; S69768
R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsot, C.
submitted to the EMBL Data Library, February 1995
A:Description: Characterization of sepa, the major extracellular protein of Shigella flexneri
A:Reference number: S57664
A:Accession: S57664
A:Molecule type: DNA
A:Residues: 1-1366 <BEN>
A:CROSS-references: EMBL:Z48219; NID:9886952; PIDN:CA88252.1; PID:9886953
R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsot, C.
Mol. Microbiol. 17, 123-135, 1995
A:Title: Sepa, the major extracellular protein of Shigella flexneri: autonomous secreted
A:Reference number: S69768; MUID:96020667
A:Accession: S69769
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-66:1077-1366 <BEN>
A:CROSS-references: EMBL:Z48219; NID:9886952; PIDN:CA88252.1; PID:9886953
A:Experimental source: strain M90T (serotype 5)
A:Accession: S69768
A:Molecule type: protein

```

```

A:Residues: 57-61,'O',63-72;'X',544-557;'X',1057-1068 <BBF>
A:Note: 6-Glu, 14-Thr, 18-Gly and 26-Gln were also found
C:Genetics:
A:Gene: sepa
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: extracellular protein
F:1-56/Domains: signal sequence #status predicted <SIG>
F:57-1366/Product: IGA-specific metalloendopeptidase homolog sepa #status predicted <
alignment_scores:
      Quality: 1003.00      Length: 1561
      Ratio: 1.263      Gaps: 60
      Percent Similarity: 50.865      Percent Identity: 25.881
alignment_block:
US-09-303-518d-651 x S57664 ..
Align seg 1/1 to: S57664 from: 1 to: 1366
22 ACACCGAAGAACACCGCAAGCCCGAAGACCGCGCGCTGCTC 71
   ||| :::::::::::::: ||| ::::|
27 ThrCysLysSerHisArg.....ArgLeuSe 35
72 GCCTGCTTACTTAGCCATATGCTGTGCTGGCATTTCTCCCAAGCTT 121
   | ::::| ::::| ||| ::::::::::::::
35 rArgArgValAlaLeuThrSerValAlaLeuSerSerSerAlaT 52
122 GG.....CGCGGACACACTTATTTGGCGATCAACCAATACATAT 162
   || :::::::::::::: ||| ||||| ::::|
52 rProAlaLeuSerAlaThrValSerAlaGluLeuProTyrGlnLeuPhe 68
163 CGCGACTTTGCCGAATAATGAAGCAAGTTGCAAGTGGCGGCAAGATAT 212
   ||| :::::::::::::: ||| ::::::::::::::
69 ArgAspPheAlaGluAsnLysGlnPheThrProGlyThrThrAsn 85
213 TGAAGTTTACACAAAGAGGAGTGTGGCGCAATATCAATGACAAAG 262
   :::::::::::::: :::::::::::::: ::::| |||
85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys..LeuAspLysA 101
263 CCGCGATGATGATTTTCTGTGTGCGGTAC..... 297
   ||||| ||||| ::::| ::::|
101 lProMetAlaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117
298 ...GGCGTGGCGCATTTGTGGCGCATATATATTTGACGCTGGCAC 344
   || ::::| ::::| ::::| ::::| ::::|
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValThrAlaLysH 134
345 TAACGGCGCTATACACAGCTGATTTGGTGGCGAAGCAAGAAATCCG 394
   ::::| ::::| ::::| ::::|
134 sValSerGlySerAspThrMetSerPheGly..... 144
395 ATGAGCACCGTTTTCTTACCAATTTGTGAAGAAATATATATTAAGCT 444
   :::::::::::::: TyraLalysAsnThrTyrThrAla 152
145 ..... 152
445 GACATTCACACCCCTTACAAGCGC...GATTACATATGCCGCTTGCA 491
   :::::::::::::: ||| |||||
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
492 TAAATTGTACAGATGCAACACCTGCGCAAGAGACAGACATGAGG 540
   ||| :::::::::::::: ||| ::::::::::::::
169 rLysLeuValThrGlyValAlaProAlaGluVal...SerAspIleGlyA 185
541 .....GGAAATACCTATTCGATTAAGAAATAATCCGACGCTGCCGC 585
   :::::::::::::: ::::| |||
185 lValAlaSerGlyAlaTyrGlnAlaGlyLysArgPheThrGluPheTyrArg 201
586 ATCGGCTCAGACACCACTATTTGGCTTATGATGATGACAAACAGGCG 635
   :::::::::::::: ::::| ::::|
202 LeuGlyGlyGlyMetGlnTyrValLys.....AspLysAsnGlyAs 215
636 TTTATCTTAC.....TCGCGGCGATGTTATTTGGCGGCAATACAC 676

```

215 nAgtHrGlnValIyTrHrThAsnGlyClyPheLeuValIcIyGlyThrValS 232
677 AATATGCAAGGTTGGGAAATAATGCG.....GTAAATTAGTTGACCGCG 720
232 eAlAlaLeuAsnSerTyRAsnAsnGlyImeTlleThrAlaGlnThrCly 248
721 GATGTGGCGCATGCGCAACGACGACTATGGCCCTATMGCGATTTGACGGTGGCG 770
249 AsPlllePheLeuIleProProAsnGlyProLeuAlaAsnTyRLeuAsnMe 265
771 AGCGCACAGCGGTTCCCAATGTTTATTTATGACAAACAAACATTAAT 820
265 TGIyAspSerGlySerProLeuPheAlaTyRAspSerLeuGlnIlySyst 282
821 GGCTGCTCAAGGAGTTTACAAACCGGCTACCTTATTCGGGACGGAA 870
282 rPvalleuIleGlyAlaLeuSerSerGlyThrAsnTyRgLy..... 295
871 AACGGTTTCCAGCTGATACGCAAGATTGGTTTCATGACATTTTCAG 920
296 AsnAsnTrPvalValIThrThrGlnAspPheLeuGlyGlnGlnProGlnAs 312
921 AGCGGATACACATACCGCTCTTTTGAACCGCGCATACGGACATTTT 970
312 nAspPheAspLysThrIle.....A 319
971 CCTTATCATCCACCAAC.....AAC 990
319 lAtyTrHrThSerGlyGlyGlyAlaLeuGlnTrpIlyTyRAspAlaAlaAsn 335
991 GGTACGGGATACGGTACACAAACCAACGAAAGGTTTCCAAATCCAAAGCT 1040
336 GLyThrIlyThrIleThrGlnGlyAsnTrHrThrTrpGly..... 348
1041 TTAAGTACAGACAGTCGCGACTGTTTGACGAATGTTGAATGAACGTGATA 1090
348 348
1091 AAGACCAAGTTTACGGCGGAGGGGTGTTAATCAGTACCGTCCAGG... 1137
349TyrAla.....TrpLysGlnAArgLys 355
1138TTAAACAACGGGTGAACCTTTCTTTATATGATTAACGGCAACGG 1181
356 LeuIleLeuAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGln 371
1182 CAACATCATGTTTCAACAACATCAACCAACGAGCGGGGGGTTTATAT 1231
371 yGlnValValIleuGlnAsnSerValAsnGlnGlyAlaGlyTyRLeuGlnP 388
1232 TTGAAGGTGATTTTACGGTCTCCCTGGAAACACACAAACGTTGGACGGC 1281
388 heAlaGlyAspTyRArgValSerAlaLeuAsnGlyGlnThrTrpMetGly 404
1282 GCGGGCGTTTCATATCAGTGAAGACAGTACCGTTACTTGAAGATTAACGG 1331
405 GLyGlyTyRleThrAspLysGlyThrHisValLeuTrpGlnValAsnGln 421
1332 CGTGGCAACAGACCGCTGTCCAAATCGGCAAGGACGCTGCACGTTG 1381
421 yAlAlaGlyAspAsnLeuHisTyRgLyGlnGlyThrLeuThValAla 438
1382 AAGCGCAAGGAGAAACCAAGGCTCGATCAGCGGTGGCGACGGTACAGTC 1431
438 snGlyThrGlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrVal 454
1432 ATTTTGGATCAGCAGCAGACGATTAAGGCAAAACCAAGCCTTTAGTGA 1481
455 lIleLeuAsnGlnGlnAlaAspArgGlyLysValGlnAlaPheSerSe 471
1482 AATCGGCTTGNTCAGCGGAGGGGTACGGTCAACGATGCGCGAATATC 1531
471 rValGlyIleAlaSerGlyArgProThrValValLeuSerAspSerGln 488
1532 AGTTCAACCCCGCAACATCTATTTGCGGTTTCGGGCGGAGTTTGAT 1581
488 lValAsnProAspAsnIleSerTrpGlyTyRArgGlyValTyRLeuGln 504
1582 TTAAACGGGCATTCGCTTTTCGTTTCCACCGCTATTTCAAAATACGAGAAAG 1631
505 LeuAsnGlyAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyRGl 521
1632 GCGGATGATTGNATCATTAATATGCGCAACACATCCACCGTTACATT. 1680
521 yAlaIleIle...ThrAsnAsnSerGlyLysLysSerThrValThrLeuA 537
1680 1680
537 sPleuGlnThrLeuLysAlaSerAspIleAsnValProValAsnThrVal 553
1680 1680
554 SerIlePheGlyGlyArgGlyAlaProGlyAspLeuTyTyRAspSerSe 570
1680 1680
570 rThIlySglnTyRpheIleLeuLysAlaSerSerTyRerProPhePheS 587
1681ACAGGCAATGAAGTATTTACACAAACGAGTGAAGATATTCAT 1725
587 eAspLeuAsnAsnSerSerValITrpGlnAsnValGlyLysAspHisAsn 603
1726 AGACTT.....AATTACAGCAAGA 1745
604 LysAlaIleAspThrValLysGlnGlnIlySileGlnAlaIleSerSeGlnP 620
1746 AATTGCTTCAACAGGTGGTTGGTGGAGAAAGATACGCAAAACGAACG 1795
620 OTyMeTyRThHisGly.....GlnLeuAsnG 629
1796 GCGGCTCAACCTGTTTACACAGCCCGCGGAGAAAGACCGCACCCNGCTG 1845
629 LysAsnMetAspValAsnIleProGlnLeuSerGlyLysAspValLeuAla 645
1846 CTTTCCGCGCGCAACAAATTTA...AACGCAACATCAACGCAACAAACGG 1892
646 LeuAspGlySerValAsnLeuProGlnIlySerIleThrLysLysSerGl 662
1893 CAACCTGTTTTCACCGCGACACGACGACCGGACGCTACATATTTAG 1942
662 TyRleuIlePheGlnGlyHisProValIleHisAlaGlyThrThrHis 679
1943 GAAGCGGTGCTGCAAAATGGAAGTATCCCAACAAGGAGAATGCTGG 1992
679 eRSeRSeGlnSer..... 683
1993 GACAACGACTGATGATNCACCGACGCTTTTAAACGGAATAATTTCCATTTCA 2042
684AspTrpGlyThrArgGlnPheThrLeuGlnIlyLysLeuLysLeuAs 698
2043 GGGCGCGCAGCGGATTTTCCCAATGATGTTCCCAAAAGTGAAGAGCGATT 2092
698 pAlaAlaThrGlnHisLeuSerAlaGlyAsn...GlyLysMetGlnIlyAspT 714
2093 GNCATTTGACCAATCAACGCCCAACGACTTTTGGTGTGCGACCGCATCAA 2142
714 lAsnAlaThrAsnGlySerThrValIleLeuGly.....Ser 726
2143 AGCCATACCAATCTGTACACGTTGGAGCTGACGACNGT...CTGACAATG 2189
727 SerArgValaIlePheTrpAspArgSerAspGlyThrGlyAsnAlaValaIleSe 743
2190 TGTGCAANAAMCAATTACCGACGATTAAGTATGATGTTCTTACTACTAGA 2239
743 rValGlnGlySerAlaThrAlaThrThrVal.....GlyAspGlnGln 757

```
2240 CNGACNTNAGCGGCGANTGTNAGNTNCCNATNACNTNNTNAAACTC 2289
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
757 eRAsPtyrSerGlyAsnValThrLeuGlnuSnlYsSerSerLeuChlIle 773
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2290 NCNGGCGTGCNNCACTNANGCAATCTTAGTCAAAATGGCGATPACAG 2339
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
774 ....MetGluArgPheThrGlyGlyLeGluAlaTyrAspSerThrIva 788
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2340 TTATACATGACGACCAACAGCCACCAAAAGCGCAACCTTAGCCTGTCG 2389
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
788 lSerValThrSerGlnAsnAlaValPheAspArgValGlySerPheVala 805
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2390 GCAATGCCCAACCAACATTTATATCAAGCCACATTAAGCGCAACNCATCG 2439
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
805 snSerSerLeuThrLeuGlyLysGlyAlaLysLeuThrAlaGlnSerGly 821
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2440 GNTTCGGCAATGCTTCATTTAATCTAAGCAACAGCCGCGCAAAACGG 2489
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
822 lIlePheSerThrGlyAlaValAlaSpValLysGlnuSnlAla..... 834
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2490 CAGTGTGACGCTTTCGACAAACGCTAACGCAAAACGTAAACCAT..... 2532
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
835 .SerLeuThrLeuThrGlyMetProSerAlaGlnLysGlnLysGlyTyrS 851
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2533 .....TCGCGACTCAAGCGCAATGTCCTCCCTAGCCGATAGGCA 2571
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
851 eRProValIleSerThrThrGlnGly...lIleSnlLeuGlnuSnlAla 866
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2572 GTATTCATTTTGAACACAGCGGCTTTCACGCAACACTACAGCGGACGCA 2621
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
867 SerPheSerValLysAsn.....MetGlyTyrIleuSerSerAsp.. 879
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2622 GGANACAGCATTAACCTTAAAGACAGCAATGACGCTGCCGTAGGCA 2671
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
880 .....lIleHsAlaGlyT 884
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2672 CGGAATTAGCAATTTAAACCTTGACAAACGCGCACCATTAACCTCATTC 2721
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
884 hTThrAlaAlaThrIleAsnLeuGlyAspSer..... 894
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2722 GCCTATCGGCAGATGCTGACAGCGCGCAACCGCAGNGTGTACAGAC 2771
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
895 .....AspAlaSpAlaGlyLysThrAspSerProLeuPheSe 907
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2772 GCGCGCGCGCGTTCGCGCCGTTCCCTATTA.....TCGCTACACGC 2815
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
907 lSerLeuMetLysGlyTyrAsnAlaValLeuArgLysIleThrGlyA 924
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2816 CAACSTCGGTAGATCCCGTTTCAACACAGCTGACGCTAAAC..... 2856
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
924 lAglInSerThrIvaIaSnMetIleAsnAlaLeuIrrTyrSerAspLys 940
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2857 .....GGCAATTTGAACNGTCAAGACATTCCTCGTTATGTGCA 2897
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
941 SerGluAlaGlyAlaLeuLysAlaLysGlySerArglLeGluLeuGlyAs 957
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2898 ACTCTTCGGCTACCGACGACAAATGAGCTGGGGGAAAGTCCGAG 2947
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
957 pGlyLysHsPheAlaThrLeuGlnValLysGluLeuSerAlaAspAsT 974
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2948 GNACTTACACCTTGGCGGTCAACAATACCGGCAACCCGTAAAGCTC 2997
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
974 hTThrPheLeuMetHsThrAsnAsnSer.....ArgAla 985
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2998 GATCAATTAAGCTAGTGAA.....GGGAAAGACAAACACCGCT 3038
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
986 AspGlnLeuAsnValThrAspLysLeuSerGlySerAsnAsnSerValle 1002
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3039 GTCCGAAACCTTAATTTCAACCTCGCAAAACGACGTCGATGCCGGCG 3088
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1002 uValAspPheLeuAsn..... 1007
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
3089 CGTGGCGTTACCACTCATCCGCAAAAGACGGCGAGTTCCGCTGCATAT 3138
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1008 .....Lys 1008
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3139 CCGGTCAAAAGACAGAGCTTCCGCAAAACCTCGGCAAGGACGAAAGCCA 3188
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1009 ProAlaSerGluMetSerValThr..... 1016
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3189 AAAACAGCGGGAAGAACACAGCGCAAGCCTTGACGCGCTGATTCGCG 3238
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1017 .....LeuIleThrA 1020
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3239 CCGGGCGCATCGCCGCAAAAGACGAAAGCGTTGCCAAGCCGCGCGG 3288
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1020 lAProLysGlySerAspGluLysThrPheThr..... 1030
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3289 CNGCAGCGCGGGGAAATGTCGGCATTAAGCAGCGCGGAGGAGAGAAAA 3338
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1031 ...AlaGlyThrGlnGlnIleGlyPheSerAsnValThrProVal..... 1044
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3339 ACGGTGACAGCGGATAAAGACAGCGCNTTGGCGAAACAGCGCGAGCG 3388
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1045 ...lIleSerThrGlnLysThrAspAlaThrLysTTP.....ValL 1058
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3389 AAACCGCGCGGNTACCACCGCTTCCCGCGCGCGCGCGCGCGCGG 3438
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1058 euthrGlyTyrGlnThrThrAlaSpAlaGlylAserLysAlaAlaLys 1074
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3439 GATTTGCCGACACCGCAGCCCAACCGCAACCTCAACCCCAACCGCAGG 3488
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1075 Asp..... 1075
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3489 CGACCTGATNAGCGGTATGCAATAGCGGTTTGAATTTTCCGCA 3538
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1076 .....PheMetAlaSerGlyTyrLysSerPheLeuThrG 1087
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3539 CCGTCAACAGCGTTTTCGCGGTACAGGACGAAATGACCGCGGTTCGC 3588
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1087 lValAlaSnAsnLeuAsnLysArgMetGlyAspLeu..... 1098
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3589 GAACACCGCGCGCAAC.....GCNGTTTGAGACAGCNGCATCCGCAAC 3632
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1099 ArgAspThrGlnGlyAspAlaGlyValTTPAlaArglIleMetAsnGlyTh 1115
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3633 CAACACTACCGTTCGCAAGATTTCCGCGCTACCGCGCAAAACGAC 3682
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1115 r.....GlySerAlaAspGlyAspTyrSerAspAsnTyrThrHisv 1129
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3683 TCGCGCAAAATCGGTATGCAAAACCTCGGACGCGCGCGCGCGCATC 3732
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1129 alGln.lIleGlyValaAspArgLysHsGlnLeuAspGlyValaSpLeu.. 1144
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3733 CTGTTTTCGCAACACCGCGCAAAACACNTTCGACGACGCGCATTCGCA 3782
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1145 .....PheThrGlyAlaLeuLeuThrTyrThrAspSerAsnAlaAs 1158
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3783 CTCGCAACGCGCTTCCCAACGCGCGCGGTTCGGGCAATACGCAATCGCA 3832
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1158 eRserHisAla.....PheSerGlyLysAsnLysSerVal 1169
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3833 CGTTCGACATCGGCATGACGACGCGCGCGGTTCATACGCGGACGCTN 3882
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1170 Gly...GlyLeuThrAlaSerAlaLeuPheAsnSerGlyAlaTyr 1184
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3883 TCAGACGCGATCGGAGCAAAATCCGCGCGCGCTGCGATTAACGCGCAT 3932
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1185 PheAspLeuIleGly.....LysTyrIleuHisHisAspAs 1196
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3933 TCAGCAGCATACGCGCGCGGTTCGCGCGATTCGCGATGCAACCGTACA 3982
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1196 gln.....HisThrAlaAsnPheAlaSerLeuGlyThrLysAspTyrS 1211
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3983 TCGCGGCAACGCGCATTTCTCCAAAAGGCGATTCACCGCTAC..... 4026
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```



```

1211  eSerHisSerTrpTyrAlaGlyAlaGluValGlyTyrAlaGlyHisLeu 1227
      ::::: ||| ::::: |||||
4027  ..... GAAACGTCATATACGCCAC 4046
      ||| ||| :::::
1228  ThrLysGluSerTrpValGluProGlnIleGluLeuValTyrGlySerVal 1244
      ::::: ||| :::::
4047  CCCCGETCTGGCTTACACCGTATACCGCGGCATTAAAGCAGATTATT 4096
      ||| ||||| |||
1244  IserGlyLysAlaPheSerTrpGluAlaArgIlyMetAla.....Ileu 1259
      ::::: ||||| :::::
4097  CATTCAACCGCGCGCAACCAINTCCATCACMCCTTATTAGCCTGTC 4146
      ||::||| ::::: :::::
1259  erMetLysAspLysAspTyrAsnProLeuIleGlyArgTrpGlyValAsp 1275
      ::::: ::::: :::::
4147  TATACCGATGCGCGCTCTGGGC.....AAGTCGCAACAGCG..... 4182
      ||| ||||| |||||
1276  ValGlyArgAlaPheSerTrpLysAspTrpLysIleHrAlaGlyAlaG 1292
      ..... GTCATACCGCGCNGTATATGGCTCAG 4207
      ..... ::::: |||||
1292  yLeuGlyTyrGlnPheAspLeuLeuAlaAsnGlyGlnHrValLeuGlnA 1309
      ::::: ||| :::::
1309  spAlaSerGlyGluLysArgPheGluGlyGlyLysAspSerArgMetLeu 1325
      || ::::: |||
4237  ..... GCGGTAACGCGCAATCAAA 4257
      ::::: ||||| |||||
1326  MetThrValGlyMetAsnAlaGluIleLys 1335

```

seq_documentation.block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - *Neisseria meningitidis* (Iga)
C:Species: *Neisseria meningitidis*
A:Valley: NG117
C:date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61319
R:Rimholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria*
A:Reference number: S61314; MUID:95302961
A:Accession: S61319
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-508 <LOM>
A:Cross-references: EMBL:X82483; NID:g732881; PIDN:CAA57866.1; PID:g732882
C:Superfamily: Iga specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```
alignment_scores:
  Quality: 870.00
  Ratio: 2.451
  Percent Similarity: 64.195
  Length: 553
  Gaps: 17
  Percent Identity: 36.890
```

```
alignment_block:
```

US-09-303-518D-651 x S61319 . .

Align seg 1/1 to: S61319 from: 1 to: 508

```

202 GCGAAATATTATAGGNTACAAACAAAAGGAGAGTTGTCGCCAATC 251
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  1 AlathirspIeuserValIlylsansIynsIngIingIyaInaInIlegIyaInaI 17

252 AATGACAAAGCCCCCATGATGATTTCTGTGTGTGGCGGTAC...G 298
    ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
  17 alenSerAsnValIPIrometIleIspheSerValAlaIaspyValaInIysa 34

299 GCGTGGCGGCGATTGGGCGGCATCATATATGTGCAGCGTGGACAT... 345
    ::|||::|||::||| ||||| ||||| ||||| ||||| |||||
  34 rglIlealathIrrValIasIapIrogIInIyIalValaIseValIyIshIaIa 50

```

[illegible]


```

885 GATACGCAAGATTGGTCTACGATGACATTAC..... 918
254 ....LysLysSerTrpGlnGluTrpSerIleTyrLysGluPheAla 269
919 .....AGAGCGATACACATACCGCTCTTTTGAACCGCGC 954
269 sPlsIleLysGlnArgAspAsnAlaGlyThrIle.....Lys 281
955 AGTAACGGA.....CATTTTCCTTACATCCACACAAAGGTAACGG 998
282 GlysAsnGlyGluHisIleTrpSerIleThrPheGly..... 293
999 TACGTAACGGAAGCAACCAAGGTTNCCAACTTAAGTAC 1048
294 .....ThrsSerHisIleGlySer.....T 301
1049 AGCAATCCGACTGTTTGACGATCTTGAATGAAGTAAGAACCA 1098
301 hAlaValAlaTrpGlu.....AlaGlyAsnGluArgAsp..... 311
1099 GTTACGCGCAGGGGTTGTAATACGTCACGCTTAACACGG 1148
312 .....AlaAsnAsnGlu 315
1149 TGAACCTTCTTTATCGATTACGGCAACGCAACTCATCTATGAA 1198
315 yGlnAsnValThrPheGluAsp.....AsnGlyThrIleValLeuAspG 330
1199 ACAACATCAACGAGCGGGCGGCTTGTATTGAAAGTATTTCAG 1248
330 InAsnIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThr 346
1249 GTCTGCGCTGAAACACGAA...ACGTGGCAAGCGCGGCGCTCATAT 1295
347 ValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspVa 363
1296 CAGTGAACGACGTACCGCTTACTGAAAGTAACGCGCTGGCAACGAC 1345
363 lAlaAspGlyLysLysValAlaTrpGluValLysAsnProAsnGlyAspA 380
1346 GCCTGTCGCAAAATGGCAAGGCAAGCGTCGACGTTCAAGCCAAAGGCA 1395
380 rGluAlaIleLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyAl 396
1396 AACCAAGGCTCGATCAGCGTGGCGACGTCAGTCATTTGGATCAGCA 1445
397 AsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLys 413
1446 GGCACACGATAAAGCAAAAAACAGCCTTAAAGTAACGCGCTGATCA 1495
413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430
1496 GCGGAGGGGTAGCGTGCACATGATGCCGATATCACTTAACCCGAC 1545
430 eGlyArgGlyThrLeuValLeuAsnSerAsnGlnIleAsnProAsp 446
1546 AAACCTATTTTCGCTTTCGGCGGCGAGCTTGGATTAAACGGGCAATC 1595
447 AsnLeuTyrPheGlyPheArgGlyLysArgLeuAspAlaAsnGlyAsnAs 463
1596 GCTTCGTTCCACCGTATTAACATACCGATGAAGGGCGATGTGNCN 1645
463 pLeuThrPheGlnHisIleArgAsnValAspGluGlyAlaArgGlyLeVala 480
1646 ATCATATGCGCAACACATCCACCGCTTACATTAACGGAATGAAGT 1695
480 snHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerLeu 496
1696 ATTACACAAACGAGTGTGAATATATCAAT 1725
497 lIleThrAspProLysThrIleSerIleHis 506

```

```

seq_name: p1r2:S61320
seq_documentation_block:
  Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (f
  C:Species: Neisseria meningitidis
  A:Variety: NGC65
  C:date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
  R:Lomholt, H.; Poulsen, K.; Mogens, K.
  Mol. Microbiol. 15, 495-506, 1995
  A:title: Comparative characterization of the iga gene encoding Iga1 protease in Neis
  A:Reference number: S61314; MIMD:95302961
  A:Accession: S61320
  A:Status: preliminary; nucleic acid sequence not shown
  A:Molecule type: DNA
  A:Residues: 1-508 <LOM>
  A:Cross-references: EMBL:X82484; NID:q732871; PIDN:CA57867.1; PID:q732872
  C:Superfamily: Iga-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloprotease

```

```

alignment_scores:
  quality: 855.00      length: 559
  Ratio: 2.395        Gaps: 19
  Percent Similarity: 63.864      Percent Identity: 36.494

```

```

alignment_block:
  US-09-303-518D-651 x S61320 ..

```

```

Align seg 1/1 to: S61320 from: 1 to: 508

```

```

202 GCGAAGATATTGAGTNTACACAAAAGGGAGTGTGGCGCAATC 251
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAACCCCGATGATGATTTTCTGTGGTCCGTAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGGCTGTTGGCGGATCAATATATTTGACCGTGCACAT... 345
34 rGluAlaThrValAlaAspProGlnIleValAlaIleSerValLysHisAla 50
346 .....AACGGCGCTATATACAGCTGATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67
372 TGGTGGGGAAGGAAGNAATCCGATCAGCACCGTTTCTTACCAATTG 421
67 pValAlaAspLysGluAsn.....GluTyrArgValY 78
422 TGAAGAACAAATATTTATACCTGCACATTCACACCTTAAACAGGC... 468
78 aGlnGlnAsnAsnTyrGluProHisLysAlaThrPheSerAlaSerAsnLeu 94
469 .....GATTNCAATATGCGCGCTTGCATTAATTTGTCCAGCA 506
95 GlYArgLeuGluLysPtyrAsnMetAlaArgPheAsnLysPheValThrG 111
507 TGCAGACCTGTGCAATGACGAGTACATGAGAGGG...AATACCTATT 553
111 uValAlaProIleAlaProThrAspAlaGlyGlyGlyLeuAspThrTyrL 128
554 CCGATTAAGAAAATATCCGAGCGTGTCCGATCGGCTCAGGACACAC 603
128 yAspLysAsnArgPheSerPheValArgValGlyAlaGlyArgGln 144
604 TAT.....TGGCGTTATGATGATGACAAACAGCGC... 633
145 LeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAsnGluLysGlyTyr 161
634 .....GATTATCTTACTCCGCGCATGATGTTAATTTGGCGGC... 669

```



```

84  ronslnysasntrphishisgllyasnngllyargleuclnlyasptryrshn 100
    ||:|||||: ||| :||| ||| :||
    ||:|||||: ||| :||| ||| :||
478 ATGCCGCGTTGGATTAATTTGGTACAGATGCGAAACCTGCGAATATGAC 527
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 MetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaProth 117
528 GAGTCAGCAGAGGGG...AATACCTATTCCGATAAGAAAAATATCCGC 574
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 rserlaclaglygllyValGluThrTyrLysAspLysAsnArgPheSerG 134
575 AGCGTCCGATCGGCTCAGACACACTATTCGCGTTATGATATGAC 624
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 lnphevalargvalGlyAlaGlyThrGlnPheGluTyrAsnSerArgTyr 150
625 AAACAGCGGATTTATCTTACTCTCCGCGCATTAATTGGCGGCAATAC 674
    ||:|||||: ||| :||| ||| :||| ||| :||| ||| :||| |||
151 AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThr 167
675 ACATATGACG.....G 685
    ||:|||||: ||| :||| ||| :||| ||| :||| ||| :||| |||
167 ofyrglnaspvalAsnValThrSerAsnLeuAsnngllyGluLeuIleG 184
686 GTTGGGGAATATATGCGGTANTTACTTTGAGCGGCGATGCGCATGCC 735
    ||:|||||: ||| :||| ||| :||| ||| :||| ||| :||| |||
184 lyphegllyaspasnseryshishisSerProglulysLeuylsGluVal 200
736 AACGACTATGGCCCTATGCCGATGCGGCGGCGGCGGCGGCGGCGGTC 785
    ||:|||||: ||| :||| ||| :||| ||| :||| ||| :||| |||
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValGluGlyAspSerGlySe 217
786 GCCAATGTTTATTTATGACAAAACAACAAATATGCTGCTCAACGAG 835
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 rProleuphealatyAspLysGlnGluAsyArgTyrPalpheueuGlyA 234
836 TTTTCA.....ACGGCTACCTTATTCGCGGAGGAAACGCTTTC 879
    ||:|||||: ||| :||| ||| :||| ||| :||| ||| :||| |||
234 latyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSerTrp 246
880 CAG.....CTGATACGCAAGATGTTCTACGATGACAT...TACAG 920
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 GlnGluTrpAsnIleTyrLysGlnPheAlaAspLysIleLysGlnArg 263
921 AGCGCATACACATACCTGCTTTTGAACGCGGCGAGTAACGCA.....C 964
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 gaspsaslaGlyThrIle.....LyselyasnngllyGluHisH 276
965 ATTTTCTTACATCCCAACAACAGCGTACGGGTACGTAACGAAC 1014
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 lstrpAsnIleThrPheGly.....Thr 283
1015 AACGAAAGGTTNTCCATCCAAAGCTTAAAGTACAGACAGTCCGACTT 1064
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 AsnSerLysIleGlySer.....ThrAlaValAlaGluLeu.. 294
1065 TGACGAATCTTGAATGAACACTGATTAAGAACAGCTTACGCGGAGGG 1114
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
295 .....AlaGlyAsnngllyLysAsp..... 300
1115 GTGTTATCAGTACCGTCCAAAGGTTAAACAGCGTGAACAACTCTCTTT 1164
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 .....AlaAsnAsnngllyGlnAsnValThrPhe 309
1165 ATCGATTACGGCAACGGCAACCTCATCTTATCAACAACATCAACAGG 1214
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 GluAsp.....AsnGlyThrLeuValLeuAspGlnAsnIleAsnngl 324
1215 CGCGGCGGCTTGTATTTTGAAGTGATTTACGCTCGCTGGAACA 1264
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 yAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyLysAsn 341
1265 ACGAA...ACGTGCAAGGCGCGGCGTTCATACAGTGAAGAACAGTACC 1311
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

341 snAspIleThrTrpLeucllyAlaGlyLysAlaValThrAspGlyLys 357
1312 GTTACTTGGAAAGTAAACGGCGTGGCAACAGCAGCGCTGTCCAAATCG 1361
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 ValValTrpGlnValLysAsnProAsnngllyAspArgLeuAlaLysIleG 374
1362 CAAGGCGCGGCTGCGATTCACAGCCAAAGGAAACAGCGGCTCGATCA 1411
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 yLysGlyThrLeucllyLeuAsnngllyThrGlyValAsnngllyGlnLeu 391
1412 GCGTGGCGGCGGTACAGTCATTTTGGATCAGCAGCGGAGATTAAGGC 1461
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
391 yValGlyLysPglYThrValIleLeuAsnngllyAlaAspAlaAspLys 407
1462 AAAAACAACGCTTTTACTGAATATGCTTGTACGCGGAGGGTACGCT 1511
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
408 LysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLe 424
1512 GCAACTGAATGCGGATATCATGTTCAACCGCGCAACCTCTATTTGGCT 1561
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 uValLeuAsnSerSerAsnngllyLeuAsnProAspAsnLeuTyrPheGly 441
1562 TTCGCGCGGACGTTTGTATTAACGCGCATTCGCTTGTCCACCGT 1611
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 heArgGlyGlyArgLeuAspAlaAsnngllyAsnAspLeuThrPheGluHis 457
1612 ATTCAAAATACCGGTGAAGGCGCATGTTGNCATATATATCCACAC 1661
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 IleArgAsnValAspGlnGlyAlaArgIleValAsnHisAsnThrSerH 474
1662 AACATCCACCGTTATCATGAGGAATGAAGATTAATACCAACGAGTG 1711
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 salaserThrIleThrLeuThrGlyLysSerLeuIleThrAsnProAsn 491
1712 GTAAGAATATCAATAGACTT 1731
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491 erLeuSerValHisSerIle 497
seq_name: p1r2:S61331
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: NG093
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61331
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-507 <LON>
A:Cross-references: EMBL:X82482; NID:q732885; PIDN:CA457865.1; PID:q732886
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease
alignment_scores:
Quality: 847.50 Length: 548
Ratio: 2.394 Gaps: 15
Percent Similarity: 64.599 Percent Identity: 36.861
alignment_block:
US-09-303-518d-651 x S61331 ..
Align seg 1/1 to: S61331 from: 1 to: 507
202 GCCAAGATATGAGGTNTTCAACAACAAAGGAGGACTGTGCGCAATC 251
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 AlatrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

```

```

252 AATGACAAAGCCCGATGATGATTTTCTGTGGTCCGGAAC...G 298
    : : : : : : : : : : : : : : : : : : : : : :
17 aleuSerAsnValrPrometileAspPheSerValAlaAspValAlaAsnLysa 34
    : : : : : : : : : : : : : : : : : : : : : :
239 GCGTGGGGCGCATTTGGTGGCGATCATATATTTGAGCCGTGGCAGAT... 345
    : : : : : : : : : : : : : : : : : : : : : :
34 rglleAlaThrValAlaAspProGlnTyrAlaValSerValLysHISAla 50
    : : : : : : : : : : : : : : : : : : : : : :
346 .....ACGGCGCATATAACAGCTGATTT 371
    : : : : : : : : : : : : : : : : : : : : : :
51 LysAlaGluValAlaHisThrPheTyrTyrGlnTyrAsnGlnHisAsnAs 67
    : : : : : : : : : : : : : : : : : : : : : :
372 TGGTGGGGAAGGAAGNATCCCGATGACACACCGTTTCTTACCAATG 421
    : : : : : : : : : : : : : : : : : : : : : :
67 pValAlaAspLysGlnAsn.....GlnTyrTrpValY 78
    : : : : : : : : : : : : : : : : : : : : : :
422 TGAAGAAGATATATTAAGCCGATGACAAATTCACACCCCTTCAACAGGC... 468
    : : : : : : : : : : : : : : : : : : : : : :
78 alGluGlnAsnAsnTyrGlnProHisLysAlaTrpGlnAlaSerAsnLeu 94
    : : : : : : : : : : : : : : : : : : : : : :
469 .....GATTANCATATGCGCGTTTGCATAAATTTGCACAGA 506
    : : : : : : : : : : : : : : : : : : : : : :
95 GlnTyrGluGlnAspTyrAsnMetalAlaArgPheAsnLysPheValThrGln 111
    : : : : : : : : : : : : : : : : : : : : : :
507 TGCAGAACCTGTGCAATGACGAGTGCACATGAGGGG...AATACCATT 553
    : : : : : : : : : : : : : : : : : : : : : :
111 cValAlaProIleAlaProThrAspAlaGlnGlyGlnLeuAspThrTyrL 128
    : : : : : : : : : : : : : : : : : : : : : :
554 CCGATTAAGAAATATCCGCGATCCGATCCGATCCGATCCGATCCGATCCG 603
    : : : : : : : : : : : : : : : : : : : : : :
128 ysaAspLysAsnArgPheSerSerPheValArgValGlnGlyAlaGln 144
    : : : : : : : : : : : : : : : : : : : : : :
604 TAT.....TGGCGTTATGATGATGACAAACAGGC... 633
    : : : : : : : : : : : : : : : : : : : : : :
145 leuValTyrGlnLysGlnAlaTyrHisGlnGlnGlnLysGlnLysGlnTyr 161
    : : : : : : : : : : : : : : : : : : : : : :
634 .....GATTATCTCTACTCCGGCGCATGGTAAATTTGGCGGC... 669
    : : : : : : : : : : : : : : : : : : : : : :
161 rAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleAlaGlnTyr 178
    : : : : : : : : : : : : : : : : : : : : : :
670 .....AATACCAT...ATGCAG 684
    : : : : : : : : : : : : : : : : : : : : : :
178 rOTyLysAspLysAlaAsnLysGlnThrMetAsnThrGlnGlnLeu 194
    : : : : : : : : : : : : : : : : : : : : : :
685 GGTGGGGAATATGCGGTANTTATGTTGACGGCGATGTCGCCATGC 734
    : : : : : : : : : : : : : : : : : : : : : :
195 GlnPheGlnAsnHisAsnLysGlnTyrSerAlaGlnGlnLeuLysGlnAl 211
    : : : : : : : : : : : : : : : : : : : : : :
735 CAACGACTATGGCCCTATGCCGATGCGAGTGGCGGACGACGCGGT 784
    : : : : : : : : : : : : : : : : : : : : : :
211 aleuSerGlnAspAlaIleuThrAsnTyrGlnValLeuGlnLysPheSerGln 228
    : : : : : : : : : : : : : : : : : : : : : :
785 CGCCAAATGTTATTTATGACAAACAAATGATGATGATGATGATGATGAT 834
    : : : : : : : : : : : : : : : : : : : : : :
228 erProLeuPheAlaPheAspLysGlnLysAsnGlnTyrPheValPheLeuGln 244
    : : : : : : : : : : : : : : : : : : : : : :
835 GTTTTACAAACCGCTACCTTATCCGCGAGGGAACCGGTTTCCAGCT 884
    : : : : : : : : : : : : : : : : : : : : : :
245 ThrTyrAspTyrTrpAlaGlnTyrGlnLysLysSerTrpGlnLysTrpAs 261
    : : : : : : : : : : : : : : : : : : : : : :
885 GATACGCAAGATGTTGTTCTGATGATGATGATGATGATGATGATGATGAT 934
    : : : : : : : : : : : : : : : : : : : : : :
261 nleTyrLysLysGlnPheAlaAspGlnLysGlnLysGlnLysGlnLysGln 277
    : : : : : : : : : : : : : : : : : : : : : :
935 CCGTCTNTTTGACCGCGGATGACGATGATGATGATGATGATGATGATGAT 984
    : : : : : : : : : : : : : : : : : : : : : :
277 laGlnTyrLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 293
    : : : : : : : : : : : : : : : : : : : : : :
985 AACACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1034
    : : : : : : : : : : : : : : : : : : : : : :
294 AsnSerHisIleGlnSerThr..... 300
    : : : : : : : : : : : : : : : : : : : : : :
1035 AAGCTTAAAGTACAGACAGTCCGACTGTTTGCAGCATCTTTGAATGAAA 1084

```

```

301 .....AlaValAlaGlnAlaAsn.....AsnGlnAla 309
    : : : : : : : : : : : : : : : : : : : : : :
1085 CTGATTAAGAACCACTTAAAGCGGCGGCGGCTTAAATCAGTACCGTCA 1134
    : : : : : : : : : : : : : : : : : : : : : :
309 rGAsp..... 310
    : : : : : : : : : : : : : : : : : : : : : :
1135 AGTTAAACACGCGTGAACACCTTTCTTTATTCATTCACGACGACGCA 1184
    : : : : : : : : : : : : : : : : : : : : : :
311 ...AlaAsnGlnGlnAsnValThrPhe.....GlnAsnGlnTyr 324
    : : : : : : : : : : : : : : : : : : : : : :
1185 ACTCATCTTATCAACATCAACACGACGCGGCGGCTTATGATGATGATG 1234
    : : : : : : : : : : : : : : : : : : : : : :
324 rLeuValLeuAspGlnAsnLysGlnLysGlnLysGlnLysGlnLysGln 341
    : : : : : : : : : : : : : : : : : : : : : :
1235 AAGGTGATTTTACGCTGCTCCGCTGAAACACAGAA...ACGTGCAAGGC 1281
    : : : : : : : : : : : : : : : : : : : : : :
341 ysaGlyAspTyrThrValLysGlnAlaAsnAsnAspIleThrTrpLeuGln 357
    : : : : : : : : : : : : : : : : : : : : : :
1282 GCGGCGCTTCATATCAGTGAACAGACGATACCTTACTTGAAAGTAAAGC 1331
    : : : : : : : : : : : : : : : : : : : : : :
358 AlaGlnLysAspValAlaAspGlnLysLysValValTrpGlnValLysAs 374
    : : : : : : : : : : : : : : : : : : : : : :
1332 CGTGGCAACGACCGCGCTCCCAATCGCAAGGACGACGTCAGCTTC 1381
    : : : : : : : : : : : : : : : : : : : : : :
374 nProAsnGlnAspArgLeuAlaLysIleGlnLysGlnTyrLeuGlnLys 391
    : : : : : : : : : : : : : : : : : : : : : :
391 snGlyThrGlnValAlaAsnGlnGlnLysGlnLysGlnLysGlnLysGln 407
    : : : : : : : : : : : : : : : : : : : : : :
1432 ATTTTGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1481
    : : : : : : : : : : : : : : : : : : : : : :
408 rLeuAsnGlnLysAlaAspAlaAspLysLysValGlnAlaPheSerGln 424
    : : : : : : : : : : : : : : : : : : : : : :
1482 AATCGCTGTGNTCAGCGCGGCGGCTGACGCTGACATGATGCGATATC 1531
    : : : : : : : : : : : : : : : : : : : : : :
424 nValGlnLysAlaSerGlnArgArgThrLeuValLeuAsnSerSerAsnG 441
    : : : : : : : : : : : : : : : : : : : : : :
1532 AGTTCAACCCGCAACATCTATTTGCGCTTTCGCGCGGACGTTTGAT 1581
    : : : : : : : : : : : : : : : : : : : : : :
441 rIleAsnProAspAsnLeuTyrPheGlnPheArgGlnLysArgLeuAsp 457
    : : : : : : : : : : : : : : : : : : : : : :
1582 TTTAAGCGGATCGCTTCCGCTCCGATGATGATGATGATGATGATGATG 1631
    : : : : : : : : : : : : : : : : : : : : : :
458 AlaAsnGlnLysAsnLeuThrPheGlnHisTyrAlaGlnValAspGln 474
    : : : : : : : : : : : : : : : : : : : : : :
1632 GCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1681
    : : : : : : : : : : : : : : : : : : : : : :
474 yAlaArgIleValAlaAsnHisAsnThrGlnHisAlaSerThrIleThrLeu 491
    : : : : : : : : : : : : : : : : : : : : : :
1682 CAGGGAATGAAAGTATTACACACGAGTGGTAAAGATATCAAT 1725
    : : : : : : : : : : : : : : : : : : : : : :
491 rGlnLysSerLeuIleThrAspProLysThrIleSerIleHis 505
    : : : : : : : : : : : : : : : : : : : : : :
seq_name: p1r2:S61330
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: SM1166
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61330
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neis
A:Reference number: S61314; MUID:95302961
A:Accession: S61330
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-508 <LOW>
A:Cross-references: EMBL:X82486; NID:q732860; PIDN:CA57869.1; PID:q732861
C:Superfamily: Iga-specific metalloendopeptidase

```

C:Keywords: hydrolase; metalloproteinase

alignment_scores:

Quality: 844.00 Length: 560
Ratio: 2.364 Gaps: 19
Percent Similarity: 63.750 Percent Identity: 36.071

alignment_block:
US-09-303-518D-651 x S61330 ..

Align seg 1/1 to: S61330 from: 1 to: 508

```

202 GCGAAGATATTGAGTNTACACAAAAAGGAGTGTGGCAATC 251
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 AlaThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17
252 AATGACAAAGCCCGCATGATTGTTTGTGTGGTGTGGCTAAC..G 298
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 AleuSerAsnValPrometIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGCATGTGGCGCATCAATATTTGACGGTGGCAT... 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 rGlleAlaThrValValAspProGlnTyraIaValSerValLysHisAla 50
346 .....AACGGCGCTATACACAGTTGATT 371
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 LysAlaGluValHisThrPheTyTyrGlnTyraGlnLysAsnAs 67
372 TGTGTCGAGAGAGAAAGNAATCCGATCAGCACCGTTTCTTACCAATG 421
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 pValAlaAspLysGluAsn.....GluTyrTrpVal 78
422 TGAAGAAGATATATTAAGCTGCATTCACACCCCTTACAGGC... 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78 AlGluGlnAsnAsnTyGlnProHisLysAlaTrpSerAlaSerAsnLeu 94
469 .....GATTATCCATATGCGCGGTTTGCATAATTTGTCACAGA 506
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GlYArGLeuGluAspTyraSnmMetaIaArpHeAsnLysPheValThrGl 111
507 TGCAGACCTGTGCAATGACGAGTGCATAGAGGG...AATACCTATT 553
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 uValAlaProIleAlaProThrAspAlaGlyGlyLeuAspThrTyrL 128
554 CGGATGAAGAAATATCCGCGAGTCCGATCGGCTCAGGACACAC 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 ysAspLysAsnArpPheSerPheValArGValGlyAlaGlyArGln 144
604 TAT.....TGCGTATGATGATGACAAACACAGGC... 633
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 LeuValTyrgLulysGlyAlaTyrlHisGlnGlnLysGlnLysGlyTy 161
634 .....GATTATCCCTACTCCGCGCATGTTAATGGCGGC... 669
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 rAspLeuArGAspLeuSerGlnAlaTyraGlyAlaIleAlaGlyThr 178
670 .....AATACACAT..ATGCAG 684
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 rGlyLysAspLleAsnIleAspGlnThrMetaThrGlnGlyLeuIle 194
685 GGTGGGAAATATAGCGGTATTGTTGACGCGCATGTGGCGCATGC 734
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
195 GlyrPheGlyAsnHisAsnThrHisTySerAlaGlnGlnLysGlnAl 211
735 CAACACATATGGCCSTATGCGGATTCAGAGTGGCGACGACGAGTT 784
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 AleuSerGlnAspAlaLeuThrAsnTyrgLylValLeuGlyAspSerGly 228
785 GCGCAATGTTATTATGACAAACAAACATTAATGGCTGCTCAACGA 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228 eTrProLeuPheAlaPheAspLysGlnLysAsnGlnTrpValPheLeuGly 244

```

```

835 GTTTTACAAACCGGCTACCCCTTATTCGGCAGGAGAAACGGTTCCAGCT 884
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
245 ThrTyraSpTyTrpAlaGlyTyrgL..... 253
885 GATACGCAAGATTGCTTACGATGACATTAC..... 918
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 ....LysSerSerTrpGlnLutrpAsnIleTyrgLysGluPheAla 269
919 .....AGAGCGATACACATACGCTGTTTGTGACCGCGC 954
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
269 sPlYsIleLysGlnArGAspAsnAlaGlyThrIle.....Lys 281
955 AGTAACGGA.....CATTTTCCTTACATCCACACACACGCGTACGG 998
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 GlyAsnGlyLunHisThrAsnIleThrPheGly..... 293
999 TACGGTACAGAAACACAGAAAGGTTNCCATCCAAAGCTTAAGTAC 1048
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
294 .....ThrAsnSerHisIleGlySer.....T 301
1049 AGACAGTCGACGTGTTGACGATCTTGAATGAACGTGAAGACCA 1098
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 hrAlaValArGLeu.....AlaGlyAsnGlnArGAsp..... 311
1099 GTTTACGCGCAGGCGGTGTTAATCAGTACCTCCAAAGTTAAACACAG 1148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 .....AlaAsnAsnG 315
1149 TGAACACCTTCTTATTATCGATTACGGCAACGCAACATCATTTATCAA 1198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 yGlnAsnValThrPheGluAsp.....AsnGlyTrpLeuValLeuAspG 330
1199 ACAACATCAACCAAGGCGCGGCGGTTGATTGAGGCTGTTTACG 1248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 LnsnIleAsnGlnGlyAlaGlyLysLeuPhePheLysGlyAspTyTrp 346
1249 GTCTCGCCTGTAACAAACGAA...ACGTGGCAAGCGCGCGCTTCATAT 1295
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 ValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyLysAsp 363
1296 CAGTGAAGACATGACGTTACTTGGAAAGTAAACGGCGTGGCAACGACC 1345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 lAlaAspGlyLysLysValAlaTrpGlnValLysAsnProAsnGlyAspA 380
1346 GCGTGTCCAAATCGGCAAGGACGCGTGCATTCACGCAAGGAA 1395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 rGLeuAlaLysIleGlyLysGlyThrLeuGluLleAsnGlyThrGlyVal 396
1396 AACCAAGGCTCGATCAGCGTGGCGAGCGTACAGTCATTTGGATCAGA 1445
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 AsnGlnGlyLeuLysValGlyAspGlyThrValIleLeuAsnGlnLys 413
1446 GCGACAGCATTAAGCCAAAAACAGCCTTTATGAAATCGCTTGNTCA 1495
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 sAlaAspSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValA 430
1496 GCGCAGGCGTACGCGTCAACATGCAATGCCGATATGCTTCAACCCGAC 1545
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 rGlyArGlyGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAsn 446
1546 AAACCTATTTCGCTTCCGCGCGGACGCTTGGATTTAACGGCGCATTC 1595
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
447 AsnLeuTyrlPheGlyrPheArGlyGlyArGLeuAspAlaAsnLysAsn 463
1596 GCTTTCGTTCCACCCTATTCAAAATACCCGATGAAGGCGCATATGNCN 1645
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 pLeuThrPheGlnHisIleArGAsnValAspGlnGlyAlaArGlyIleValA 480
1646 ATCATATGCGCACACACATCCACCGTTTCACTTACAGGAGTAAGT 1695
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
480 snHisAsnThrAspArGAlaSerThrIleThrLeuThrGlyLysSerLeu 496
1696 ATTACACACCGAGTGTGAAGATATCAAT 1725

```


1074 TTGATGAACTGATTAAGAACAGATTACGGCAGGGGTTTAATC 1123
 294AlaGlyasn..... 296
 1124 ACTACGTCACAGGTTAAACAAGGTGAACCTTCTTTATTCGATTAC 1173
 297GluThrGlyAlaAsnAsnGlyGlnAsnValThrPhe.....Glu 309
 1174 GGCACAGCGCAACATCATCTTATCAACAACATCAACCAAGCGCGGGG 1223
 310 AsnAsnGlyThrLeuValLeuAspGlnAsnGlnGlyAlaGly 326
 1224 TTTGATTGTAAGGTGATTTACCGTCGCTGCAAAACAAGCA...A 1270
 326 yLeuPhePheLeuGlyAspPheThrValLeuGlyLeuAsnAsnAspPhe 343
 1271 CGTGGCAAGCGCGCGGCTTCATTCAGTAAGACAGTACCGTTACTGG 1320
 343 ThrThrLeuGlyAlaGlyLeuAspValAlaAspGlyLysLysValValTrp 359
 1321 AATGTAACAGCGCGTGCACACAGCCGCTGTCACAAATCCGCAAGGAC 1370
 360 GlnValAlaAsnProAsnGlyAspThrLeuAlaLysIleGlyLysGlyTh 376
 1371 GGTGACGTTCAAGCCAAAGGGAAACCAAGCTCGATCAGCGTGGCG 1420
 376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGly 393
 1421 ACGGTACAGTCAATTTGGATCAGCAGGCGACGATTAAGCAAAACAA 1470
 393 spGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGlnLysValGln 409
 1471 GCCTTAGGAATGGCTGTGTCAGCGGCGGATGCGTCACTGCACTGA 1520
 410 AlaPheSerGlnValIleValIleValSerGlyThrGlyThrLeuValLeuAs 426
 1521 TGCCGATTAATCAGTTCACCCGACAACTATTTGCGCTTCGCGAGC 1570
 426 nSerSerAsnGlnIleAsnProAsnLeuThrPheGlyPheArgGlyG 443
 1571 GACGTTGATTTAAACGGCATTCGCTTCGTCACCGTATTCAAAT 1620
 443 LysrLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleLeuArgAsn 459
 1621 ACCGATGAAGGGCGGCGATGTCNCNTACTAATGCGACAAACATCCAC 1670
 460 ValAspGlnGlyAlaArgIleValAsnHisAsnThrGlyHisThrSerTh 476
 1671 CGTTACCATTAAGGAATGAAGTATTACACACCGAGTGTAAAGATA 1720
 476 rIleThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerV 493
 1721 TCATAGACTT 1731
 493 496
 493 aHisSerIle 496

seq_name: p1r2:S61327

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm

C:Species: Neisseria meningitidis

A:Variety: NC690

C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S61327

R:Lombolt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseria

A:Reference number: S61314; MUID:95302961

A:Accession: S61327

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOM>

A:Cross-references: EMBL:X82479; NID:9732883; PIDN:CA57862.1; PID:9732884

C:Superfamily: IgA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 840.00 Length: 538
 Ratio: 2.360 Gaps: 17
 Percent Similarity: 66.171 Percent Identity: 37.732

alignment_block:
 US-09-303-518D-651 x S61327 ..

Align seg 1/1 to: S61327 from: 1 to: 496

202 GCGAAGATATGAGTNTACACAAAGGGAGTGGTGGCAATC 251
 1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleLysAl 17
 252 AATGCAAAAGCCCGATGATGATTTCTGTGTGTCGCGTAAAC...G 298
 17 AlaSerAsnValProMetIleAspPheSerValAlaAspValAsnArg 34
 299 GCGTGGCGCATTTGGTGGCGATCAATATTTGTGACCGTGCACATAC 348
 34 rGThrLeuThrValIleAspProGlnThrAlaValSerValLysHisVal 50
 349 GCGGCTATTAACAACGTTGATTTGTGCGGAGGAAGAAATCCGAT... 396
 51 LysGlyAspGlnIleSerTyrTyrGlyHisAsnGlyHisLeuAspVa 67
 397 ...CACACCGCTTTCTTACCAATTTGTAAAGAAATATATTAAGC 442
 67 LSerAsnAspLysAsnGlnLysThrGlySerValAlaGlnAsnAspLysGlu 84
 443 CTGCAATCA...CACCTTACAGCC.....GATTNCA 477
 84 rAsnLysAsnThrPheHisIleGlyAsnGlnGlyThrLeuGluAspTyrAsn 100
 478 ATGCGCGCTTTCATTAATTTGTCACAGATGACAGAACCTGTCGAATGAC 527
 101 MetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaProth 117
 528 GAGTGACATGAGGGG...AATACCTTATCCGATTAAGAAATATCCG 574
 117 rSerAlaGlyGlyGlyAlaGluThrTyrLysAspLysAsnArgPheSerG 134
 575 AGCGTGCAGCATGCGTCAGGACACACCATATGATGATGATGAC 624
 134 IupheValArgValGlyAlaGlyThrGlnPheGluTyrAsnSerArgTyr 150
 625 AATACAGCGCATTTATCTTACTCCGCGCATGTTAATTTGGCGCAATAC 674
 151 AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrPr 167
 675 ACATATGCAG.....G 685
 167 oTyrGlnAspValAlaAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
 686 GTTGGGGAATATGCGGTANTTAGTTGAGCGCGCATGTGCGGCATGCC 735
 184 LysPheGlyAspAsnSerLysHisHisSerProGlnLysLeuLysGlyVal 200
 736 AACGACTATGGCCCTATGCCGATGACAGTGGCGGACGACACACGGTTC 785
 201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuLysSerGlySer 217
 786 GCGATGTTTATTATGACAAACAAACATTAATGGCTCTCTACAGGAG 835
 217 rProLeuPheAlaTyrAspLysGlnGluLysArgTyrValPheLeuLys 234
 836 TTTTACAA.....ACCGGCTACCTTATTCGCGAGGAAGAAAGTTTC 879
 234 IatYrAspTyrTyrPalaGlyTyr.....GlnLysAsnSerTrp 246


```

167 oTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
686 GTTGGGAAATTAATGGCGTANTAGTTTGAGCGGCGATGTGGCCATGCC 735
184 LyrheGlyAspAsnSerLysHisHisSerProGlyLysLeuLysGlyVal 200
736 AACGACTATGGCCCTATGCCGATTTGAGTGGCGGACGACAGCGGCTTC 785
201 LeuSerGlnAsnAlaLeuThrAsnGlyAlaValLeuGlyAspSerGlySer 217
786 GCCAATGTTTATTAATGACAAACAAATTAATAGGCGCTGACAGCGAG 835
217 rProLeuPheAlaTyrAspLysGlnGlyLysArgTyrValPheLeuGlyA 234
836 TTTTACAA.....ACCGCTACCTTATTCGCGACGGGAAACGGGTTC 879
234 LArGAspTyrTyrPheAlaGlyTyr.....GlnLysAsnSerTyr 246
880 CAG.....CTGATACGCAAAAGATTGGTCTACGATGACATTTACAGAG 923
247 GlnGlyThrAsnIleTyrLysGlnPheAlaAspGlyLysGlnIle 263
924 CGATACACATACCGCTCTTTTGAACGCGCCAGTAAAGCAGCATTTTCTCT 973
263 GAsp...AsnAlaGlyThrIleLysGlyTyrGlnHisHisTyrLysT 279
974 TTACATCCAAACACACGCGTACGCGTACGATACAGAAACCAAGAAAG 1023
279 hrThrGlyThrAsnSerHisIleGlySerThr..... 289
1024 GTTTCGATCCAAAGCTTAAAGTACAGACGCTGCTTGTGACGATC 1073
290 .....AlaValArgLeu..... 293
1074 TTTGATGCAAACTGATTAAGAACAGATTACGCGGAGGGGTGTAATC 1123
294 .....AlaGlyAsn..... 296
1124 AGTACCGTCCAAAGTTAAACAGCGTGAACCTTTCTTTATGATTTAC 1173
297 .....GlnArgGlyAlaAsnAsnGlnGlnAsnValThrPhe.....Glu 309
1174 GGCACGCGCAAACTCATCTTATCAAAACATCAACCAAGCGCGCGCG 1223
310 AsnAsnGlyThrIleValLeuAspGlnAsnIleAsnGlnGlyAlaGly 326
1224 TTTGATTTTGAAGTGAATTTTACGCTCCGCTGCAAAACCAAGAA...A 1270
326 YLeuPhePheLysGlyAspTyrThrValLysGlyIleAsnAsnAspIleT 343
1271 CGTGGCAAGGCGCGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGG 1320
343 hrThrLeuGlyAlaGlyIleAspValAlaAspGlyLysValValThr 359
1321 AAAGTAAACGCGGTGCAACAGCGCTGCAAAATGCGCAAGGAC 1370
360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
1371 GGTGACGTTTCAAGCCAAAGGGAACCAAGGCTGATCAGCGGCGGCG 1420
376 rLeuGlnIleAsnGlyThrGlyValAlaAsnGlnGlyIleLeuLysVal 393
1421 ACGGTACAGTCAATTTGATCAGACGCGAGCATTAAGGCAAAACAA 1470
393 spGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGlnLysValGln 409
1471 GCCTTTAGTAAATGGCTTGTACAGCGGACGAGGCTAGCGTCAACTGAA 1520
410 AlaPheSerGlnValGlyIleValSerGlyArgArgThrLeuValLeuAs 426
1521 TGCCGATTAATCAATCCGACCAAACTATTTGCTTCCGCGGCG 1570
426 nSerSerAsnGlnIleAsnProAsnAsnLeuTyrPheGlyPheAlaGlyG 443

```

```

1571 GACGTTTGATTTAAACGCGCATTCGCTTGTCTCCACCGTATTCAAAT 1620
443 LArGLeuAspAlaAsnGlnLysAsnLysPheThrPheGlnHisIleArgAsn 459
1621 ACCGATGAAGGCGCGATGATGTCNATCATTAATGCAACACATCCAC 1670
460 ValAspGlnGlyAlaArgIleValAlaAsnHisAsnThrGlyHisThrSer 476
1671 CGTTACCATTAACGGAATGAATGATTTATACCAACCGAGTGTGAAGATA 1720
476 rIleThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSer 493
1721 TCATATGACTT 1731
493 ..... 496
493 aHisSerIle 496
seq_name: plr2:S61322

```

```

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: HF54
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61322
R:Rimbold, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the Iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61322
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82473; NID:g732877; PIDN:CAA57856.1; PID:g732878
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase, metalloproteinase

```

```

alignment_scores:
Quality: 815.00 Length: 534
Ratio: 2.309 Gaps: 15
Percent Similarity: 66.105 Percent Identity: 36.704

```

```

alignment_block:
US-09-303-518d-651 x S61322 ..

```

```

Align seg 1/1 to: S61322 from: 1 to: 496

```

```

202 GCGAAGATTTATGAGGTTTACCAACAAAGGAGGTGTCGCAAAATC 251
1 AlArThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAGCCCGATGATGATTTTCTGTGTGTCGCGGTAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAlaArgA 34
299 GCGTGGCGGATTTGGTGGCGGATCATATATGTTGTGCGGACATTAAC 348
34 rGlnThrLeuThrValIleAspProGlnTyrAlaValSerValIleHisVal 50
349 GCGCGCTATTAACAACGTTGATTTTGTGCGGAGGAAGAAATCCGAT... 396
51 LysGlyAspArgLysIleSerTyrTyrGlnHisAsnGlnLysIleAspVal 67
397 ....CACACCGTTTCTTACCAAAATTTGAAAAGCAATTAATTAAGC 442
67 LArAsnAspGlnAsnGlnTyrArgSerValAlaGlnAsnAspTyrGlnP 84
443 CTGCAATTC.....CACCGTTACAAAGCG.....GATTANCA 477
84 rAsnLysAsnThrPheHisIleGlyAsnGlnGlnLysArgLeuGlnAspTyrAsn 100
478 ATGCCGCGTTTGCAVAAATTTGTCACAGATGCAGAACCTGTGCAATGAC 527

```

```

101 MetAlaIrgLeuAsnLysPheValThrGluValAlaProIleAlaProTh 117
528 GAGTGCATGAGGGG...AATACCTATTCGGATTAAGAAATAATATCCG 574
117 rseAlaIglYlGlyValGluThrLysAspLysAsnArgPheSer 134
575 ACCGTGTCGGCATCGGCTCAGACACCACATATGGCGTTATGATGATGAC 624
134 IupHeValArgValGlyAlaGlyThrGlnPheGluIuIyrAsnSerArgTyr 150
625 AAACACGGCATTTATCTACTCCGGCGCATGTTAATGGCGGCAATAC 674
151 AsnMetHrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyHrPr 167
675 ACATATGAC...G 685
167 oTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGluGlyLeuIleG 184
686 GTTGGGGAATTAATGCGGTANTTAGTTCAGCGCGATGTGCGCATGCC 735
184 LyrHedLysAspAsnSerLysHisSerProGluLysLeuLysGluVal 200
736 AAGGACTATGGCCCTATGCCGATTCGAGGTGCGGCGACAGCGGTTTC 785
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValAlaGlyAspSerGlySe 217
786 GCCAATGTTATTATGACAAACAAACAAATAATGGCTGCACACAGGAG 835
217 rProLeuPheAlaTyrAspLysGlnGluLysArgTyrPheValPheLeuGlyA 234
836 TTTTACA...ACCGCTACCTTATTCGCGCAGGAGAAACGGTTTC 879
234 LatyrAspTyrTrpAlaGlyTyr...GlnLysAsnSerTrp 246
880 CAG...CTGATACGCAAGATTGGTCTACGATGACATTACAGAG 923
247 GlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIleLysGlnAr 263
924 CGATACACATACCGTCTNTTTGAACCGCGCAGTAAACGACATTTTTCCT 973
263 gAsp...AsnAlaGlyThrIleLysGlyTyrGluLysHisTrpLysT 279
974 TTACATCCACAAACAGGTACGGGTACGATACAGAAACCAACGAAAG 1023
279 hrThelGlyThrAsnSerHisIleGlySerThr... 289
1024 GTNTCCATCCAAAGCTTAAGTACAGACAGTCGCGAGTGTTCAGAGATC 1073
290 .....AlaValAlaGlu... 293
1074 TTTGAATGAACGTATAAAGACAGTTTACGCGGAGGGGTGTTAATC 1123
294 .....AlaGlyAsn... 296
1124 AGTACCGTCCAAAGTTAACAACAGGTGAACCTTTCTTTATGATAC 1173
297 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe...Glu 309
1174 GGCACGCGCAACCTCATCAACAACATCAACAGCGCGGCGG 1223
310 AsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGlyG 326
1224 TTTGATTTTGAAGTGAATTTAGGTTCCCTGAACAAACAGAA...A 1270
326 yLeuHePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyIleT 343
1271 CGTGCAAGGCGGCGCTCATATACATGATGAAGACAGTACGTTACTGG 1320
343 hrTrpLeuGlyAlaGlyLysValAlaAspGlyLysValValTrp 359
1321 AAAGTAAACGCGGTGGCAACAGCGCTGCAAAATCGGCAAAAGGCAC 1370

```

```

360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
1371 GCTGCACGTTTACGCCAAAGGGGAAACCAAGGCTGCATGACGCTGGCG 1420
376 rLeuGluIleAsnGlyThrGlyValAlaAsnGlnGlyLysLeuLysValGlyA 393
1421 ACCGTACAGTCTTTTGGATCGACGACGACGATAAAGCAAAACAA 1470
393 spGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLysValGln 409
1471 GCCTTTGAATATCGCTTGTACGCGCAGGCGGTACGGTGAACATCA 1520
410 AlaPheSerGlnValGlyIleValSerLysArgArgThrLeuValLeuAs 426
1521 TGCGATATATCAGTTCAACCCGCAACACTATTCGCGTTTCGCGCG 1570
426 nserProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPheArgGlyG 443
1571 GACGTTTGAATTTAAGCGGATTCGCTTTCGTTCCACCGTATTCAAAT 1620
443 LyrGluLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleArgAsn 459
1621 ACCGATGAAGGCGCATGATTCGNCATCATATGACCAACACATCCAC 1670
460 ValAspGluGlyAlaArgIleValAlaAsnHisAsnThrAspAlaSerTh 476
1671 CGTTACCTTTCACAGGAATGAAGTATTTACCAACGAGTGTGAAGATA 1720
476 rIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuSerV 493
1721 TC 1722
493 al 493

```

seq_name: pir2:S61332

seq_documentation block:

IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (str N:Alternate names: Iga1 protease
C:Species: Haemophilus influenzae
A:Variety: HK635
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria meningitidis strains
A:Reference number: S61314; MUID:95302961
A:Accession: S61332
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-462 <LOW>
A:Cross-references: EMBL:82488; NID:g732714; PIDN:CA57871.1; PID:g732715
A:Experimental source: strain HK635
A:Note: The authors did not translate the codon for residue 462
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

alignment_scores:

Quality:	760.00	Length:	510
Ratio:	2.353	Gaps:	25
Percent Similarity:	63.333	Percent Identity:	38.627

alignment_block:

US-09-303-518D-651 x S61332 ..

Align seg 1/1 to: S61332 from: 1 to: 462

```

202 GCGAAAGATTTAGGCTNTTACCAAAAGGCGAGTGTGCGCAATC 251
||| .....||| ||| .....||| .....||| .....|||
1 AlaThrAsnValGluValArgAspLysAsnHisSerLeuGlyAsnAl 17

```


alignment_scores:

Quality: 687.00 Length: 511
 Ratio: 2.134 Gaps: 24
 Percent Similarity: 63.014 Percent Identity: 35.616

alignment_block:

US-09-303-518D-651 x S61333 ..

Align seg 1/1 to: S61333 from: 1 to: 471

```

202 GCGAAGATATTTGAGTNTACACAAAGGAGGAGTGTGGCGCAATC 251
    ||| :|||||:||||| :|||||:||||| :|||||:|||||
    1 AlathrsnvalglvalargaspIysasnslnglnserleuglseral 17
252 AATGACAAAGCC...CCGATGATGATTTTCTGTGCTGCGCGTAAAC 297
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
    17 aleuproasnnglylerproketileasprheservalvalasprvalsnl 34
298 .GGCGTGGCGCATTTGGTGGCGCATATATATTGTGAGCTGGCAAT 345
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
    34 ysarglileglythrleuvalasprproglntyllevalserValysHs 50
346 AAC.....GGCGCATATACACCTTGA 368
    :|||||:|||||:|||||:|||||:|||||:|||||
    51 AlahlslyshislleasnsprneIyrpneglyHsItyasn..... 64
369 TTTTGTGGCGAAGAAATCCGATGACGACCGTTTCTTACCAAA 418
    ||| :|||:|||||:|||||:|||||:|||||:|||||
    65 ...GlyHsIargAsprvalserasprgluasnlys...TyrSerV 78
419 TTGTGAAAGAAATATATTAACCTGACATTCACACCTTACAAAC... 465
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
    78 alvalthrghlnasnsvalIysproasnnglnasprtrphIsvalasprlys 94
466 .....GGCGATTANCATATGCGCGTTCGATTAATTTGTCAAGATCG 509
    ||| :|||:|||||:|||||:|||||:|||||:|||||
    95 ArgleuAsprAsprlyrIasnmethproargleuasnIysphevalthrleuval 111
510 AGAAGCTGTGCAAAATGACAGTGCATGAGGGCAAT.....ACCT 550
    :|||||:|||||:|||||:|||||:|||||:|||||
    111 lalaprthr.....ThrprothrleualaglAsprAsprleugluthr 126
551 ATTCCGATTAAGAAATATCCGAGCGTTCGCGATCGCTGACAGAC 600
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    126 yrllysasprlysgIuIysIyrproserphevalargvalglvalaglythr 142
601 CACTAT.....TGGCGTTA 614
    :|||||:|||||:|||||:|||||:|||||:|||||
    143 glnphevaltyrgluIysglserIyryrvalgluIysthrThrargAs 159
615 TGATGATGACAAACAGCGGATTTATCCACTCCGCGCATGTTAAATTG 664
    :|||||:|||||:|||||:|||||:|||||:|||||
    159 msnasprileysrheleuAsprgluIalarg.....Argtyralileg 174
665 GCGGCATACACATATGACAGGTTGGGAATATATGCGTANTTAGTTG 714
    ||||| :|||:|||||:|||||:|||||:|||||:|||||
    174 lygl...ThrprotyrgluIyIleasnIleasprproserIasnlys 189
715 AGCGC.....GATGT 725
    :|||||:|||||:|||||:|||||:|||||:|||||
    190 lysgllyleuIleglypheglyAsprserargluasnHsvalIleaspral 206
726 GCGCCATGCGCAAGCACTATGCGCTATGCGCATTCGAGTGGCGAGCG 775
    :|||:|||||:|||||:|||||:|||||:|||||
    206 alysrthrleuAsprleuAsprproleuthrAsntyrcllyallleugly 223
776 ACAGCGGTTCCCAATGTTTATTAAGCAAAACAAATTAATGCGTG 825
    |||||:|||||:|||||:|||||:|||||:|||||:|||||
    223 spserglserproleuAsprleuAsprlysglnIasnlystrypal 239
826 CTCACAGGAGGTTTACAAACCGCGCATCCCTATTCC..... 861
    :|||:|||||:|||||:|||||:|||||:|||||
    240 pheilegly.....ProtyrthrtyrItpalaglyty 250

```

```

862 .GGCAGGAAAMC.....GGTTCCAGCTGATACGCAAAAGATGTTCT 904
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    250 rgllyIyssertrpglnIutrpasnleryrIysIysasp...Phea 266
905 AGCATGACATTTACAGAGCGGATACACATACCGTCTNTTTGACCGCG 954
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    266 laAsprsnlleIysIysargAsprasnIagluAlaval.....Prophe 280
955 AGTAACGACATTTTCTTACATTCACAAACAGGATGACGATACGTT 1004
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    281 serThrserclutyrHsItrphrAsnthrtrphasnHsIsglnsergluI 297
1005 AACAGAAACCAAGAAAGGATTCGATTCGAAAGCTTAAGTACAGAG 1054
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
    297 elysasnthrAsprHsIthrIle.....ThrV 306
1055 TCCGACTGTTTACGATCTTGAATGAACGTGAAGAACACAGTTTAC 1104
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    306 alThrleu.....ProserAsprAsnIargleu... 315
1105 GCGGACGAGGGGTGTAAT...CAGTACCGTCCAAAGTTAAACACGCTGA 1151
    ||||| ||| :|||:|||||:|||||:|||||:|||||
    316 .....ValasnphelnglnIysgluHsIleuglnthrIygl 328
1152 AACCTTCTTTTATCGAT.....TACGCAACGCGCAACTCATCT 1192
    :|||||:|||||:|||||:|||||:|||||:|||||
    328 nasnvalthrPheasprserthrAsnasngIyIysgllythrleuIlel 345
1193 TATCAAAACATCAACCAAGCGCGGTTGTATTTGAAGTGAT 1242
    |||:|||||:|||||:|||||:|||||:|||||:|||||
    345 euAsprAsprHsIleasnIlncllyalaglyleuphepelysglyasn 361
1243 TTTACGCTCTGCGTGAACACGAA...ACGTGCAAGCGCGGCGCT 1289
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
    362 tyrgluvalIysgllyIysThrAsprIlethrtrpvalgllygllyl 378
1290 TCATTCAGTGAAGACAGTACCGTTACTTGAAGTAAACGCGCTGCCAA 1339
    :|||||:|||||:|||||:|||||:|||||:|||||
    378 easprvalalaglIuIyIysgluvalIvaltrpIyvalHsIasnproglul 395
1340 AGAAGCGGCTTCCAAATCGCAAGGACGCGTTCGCGTCAAGCCCAA 1389
    :|||||:|||||:|||||:|||||:|||||:|||||
    395 ysAsprHsIleualalysIleIyIysgllythrleuIlevalglnglylys 411
1390 GGGGAAACCAAGCTCGATGACGCTGGCGAGCGTACAGTATTTGA 1439
    |||||:|||||:|||||:|||||:|||||:|||||:|||||
    412 glIysnasnIysgllyserleuIysvalglYAsprIythrvalIleuIy 428
1440 TCACAGGACAGCATTAAGCAAAACAGCGTTAGTGAATGCGGT 1489
    |||||:|||||:|||||:|||||:|||||:|||||
    428 sglnglnthrAsnIyserIy...gluHsIalphealserValgllyl 444
1490 TGNTCAGCGGAGGAGTACGCTGCACTGAATCCGTAATCACTTCAAC 1539
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
    444 levalserglYargserThrValIvalIeUasnasprAsprlysglnvalAsp 460
1540 CCCGACAACCTATTTGCGCTTTCGCGGCGGA 1572
    |||:|||||:|||||:|||||:|||||:|||||
    461 ProAsnserIleIytrpneglypneargIyglly 471

```

seq_name: p1r2:T00317

seq_documentation_block:

probable serine protease esp. extracellular - *Escherichia coli* plasmid p0157
 N:Alternate names: putative exoprotein precursor
 C:Species: *Escherichia coli*
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Dec-2000
 R:Accession: T00317; T42120
 R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
 S.; Shingawa, H.
 DNA Res. 5, 1-9, 1998
 A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemo

A:Reference number: Z14127; MUID:98290540
 A:Accession: T00317
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1300 <MAK>
 A:Cross-references: EMBL:AB011549; NID:94589740; PIDN:BA031836.1; PID:93337077
 R:Burkhardt, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
 Nucleic Acids Res. 26, 4196-4204, 1998
 A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Escher
 A:Reference number: Z22068; MUID:98391744
 A:Accession: T42120
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1300 <BUR>
 A:Cross-references: EMBL:AF074613; PIDN:AAC70088.1
 A:Experimental source: strain EDL933; serotype O157:H7
 C:Genetics:
 A:Gene: espP
 A:Genome: plasmid pO157
 C:Superfamily: 19A-specific metalloendopeptidase
 C:Keywords: serine proteinase

alignment_scores:
 Quality: 631.50 Length: 1554
 Ratio: 0.882 Gaps: 68
 Percent similarity: 46.075 Percent identity: 21.815

alignment_block:
 US-09-303-518D-651 x T00317 ..

Align seg 1/1 to: T00317 from: 1 to: 1300

```

82 TTAGCATATGCGTGTGCGTGGCATTCCTCCCAAGCTTGCGGCGGACA 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 LeuAlaLeuCySPhe...LeuGlyLeuLeuInserSerTySerPheAl 55
132 CACTATTTCGGCATCACTACCATAC.....TATGCGGACTTG 172
::: ::::|:|||||:|||||:|||||:|||||:|||||:|||||:
55 aserInlelaspIleSerasnPhetyrIleArgaspTyrmetsphea 72
173 CGAAATTAAGCAAGTTTGCAGTGGGCGGAAAGATTTGAGCTTAC 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 IagIlnsnlysglyIlePheGlnAlaGlyAlaThrAsnIleGlnLeval 88
223 AACAAAAAAGGAGTGTGTCGCAATCATGACAAAAAGCCCGCATAT 272
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
89 LysLysAspGlySerThrLeu.....LysLeuProGluValProPhePr 103
273 TGATTTTCTGTGTCGCGTAAAGCGGCGGCGGCGGCGGCGGCGG 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 oAspPheSerProValAlaAsnlysglySerThrThrserIleGlyGly 120
323 AATATTTGTGAGCGTGCATTAACGGCGGCTATTAACAGCTTGATTT 372
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
120 IatYserIleThrAlaThrHisAsn..... 128
373 GTGCGGAAGGAAGNATCCGATCAGCACCCTTTTCTTACCAATGT 422
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
129 .....ThrLysasnHisSerValAlaThrGlnAsnTr 140
423 GAAAGCAATTAATTAAGCT.....GACAATTCACACCCCTTACA 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 pGlyasnSerThrTyrlsgIlnThrAspTrpAsnThrSerHisPro.... 155
464 ACGGCGATTAACATATGCGGCTTGCAATTAATTTGTACAGATGCAGAA 513
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
156 .....AspPheAlaValSerArgLysAspLysPheValValGluThrArg 170
514 CCGTGTGAATGACGAGTGACATGAGGGGGAATACTATTCCGATTAAGA 563
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
171 GlyAlaThrGlnGlyAlaAspIleSerLeuSerLysGlnGlnAlaLeuGl 187

```

```

564 AAATATCCCGAGCGT.....GTCCGA 586
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
187 uArgTyrlGlyValAsnTyrlGlyGlyLysLeuIleAlaPheArgA 204
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
587 TCGGCTCAGACACCATTAATGGCGTTATGATGATGACAAACAGCC... 633
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
204 IagIlyserLy.....ValValSerValLysLysAsnGlyArg 216
634 .....GATTATTCCTACCGCGCGCATGGTATTAAGGCG 668
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
217 IleThrProPheAsnGluValSerTyrlLysProGluMetLeuAsnGlySe 233
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
669 CAATACCATATGACAGGTTGGGAAATATGCGCTAATTAAGTGAACG 718
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
233 rPheValHisIleAspAspTrp.....SerGlyTrpLeuIleThrA 248
719 GCGATGTGCGCCATGCCAACGACTATGGCCCTATGCCGATTCAGGTGG 768
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
248 sn.....AsnGlnPheAspGluPheAsnIleAlaSer 259
769 GCAGGCGACAGCGGTTGCGCAATGTTATTATGACAAACAAACATTA 818
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
260 GlnGlyAspSerGlySerAlaLeuPheValTyraSpasnGlnLysLys 276
819 ATGCTGTCTCAACGAGTTTAAACCGGCTACCTTATTCGCGGAGG 868
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
276 strValValAlaIagIlyThrValTrpGlyIleTyraSnTyrlAlaSnGlyL 293
869 AAACGCTTCCAGCTGATACGCAAGATTGG.....TTCTACGAT 909
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
293 ysaSn.....HisAlaIaTySerLysTrpAsnGlnThrThrIleAsp 307
910 GACATTTACAGAGCGATACATACATACCGTCTTTTGAACCGCGCATTA 959
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
308 AsnLeuLysAsnLysTyrlSerTyraSnValAlaSpmet.....SerG 321
960 CGGACATTTTCTTACATCCACACAAAGGATAGGCTAGCGGTAAACG 1009
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
321 yAlaGlnValAlaThrIleGlnAsnGlyLysLeuThrGlyThrGlySerA 338
1010 AAACCAACGAAAGGINTCCAAATCCAAAGCTTAAGTACAGACAGTCCGA 1059
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
338 spThrTrpAspIle..... 342
1060 CTGTTGACGAATCTTGATGAATGAATGATTAAGAACAGATTACCGGCG 1109
342 ..... 342
1110 AGGGGTTGTAATCAGTACCGCTCAAGTTAAACACGGTGAAGAACTTT 1159
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
343 .....LysAsnLysA 346
1160 CTTTATCATTTACGGCAACGCGCAACTCATTTATCAACACATCAAC 1209
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
346 spLeuIlePheThrIleGlyGlyAspIleLeuLeuLysSerSerPheAsp 362
1210 CAAGCGCGCGGCTTGTATTTGAAGTGATTTAAGCTTCGCGCTCA 1259
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
363 AsnGlyIaIagIlyLeuValPheAsnAspLysLysThrTyrlArgValAs 379
1260 AAACCAACGAA...ACGTGCAAGGCGGCGGCTTCATATCATAGTAAGACA 1306
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
379 nGlyAspAspPheThrPheLysGlyAlaGlyValAspThrArgAsnGlyS 396
1307 GTACCGTTACTTGAAGTAACGCGCTGCAACGACCGCTGTCCAA 1356
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
396 erThrValGluTrpAsnIleArgTyraSpasnLysAspAsnLeuHisLys 412
1357 ATGCGCAAGGCGACGCTGCACGTTCAAGCCAAAGGGAACCAAGAGCTC 1406
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
413 IleGlyAspGlyThrLeuAspValArg...LysThrGlnAsnThr...As 427

```



```

1407 GATCAGCGGTGGCGACGATCATTTTGGATCAGCAGCAGCAGCATA 1456
      ::::::::::::::::::::|
427 nleuLysThrGlyGlnGlyLeuValIleLeu..... 437
1457 AAGGCAAAAAACAAGCCTTTAGTAAATGCGCTTGNTCAGCGCAGGGGT 1506
      ||| ::::::::::::::::::::|
438 .GlyAlaGlnLysThrPheAsnAsnIleThrSerGlyAspGly 453
1507 ACGGTGCACTGAATGCCGATATCAGTTACACCCGCAAA..... 1548
      ||| ::::::::::::::::::::|
454 ThrValAlaGlyLeuAsnAlaGlnAsnAlaLeuSerGlyGlyGlnGlyAsnGln 470
1549 .CTCTATTTGGCTTCGCGCGAGCTTGATTTAAAGCGGCATTCG 1597
      ::::::::::::::::::::|
470 yllePhePheAlaLysAsnGlyLysThrLeuAsnGlyLysAsnG 487
1598 TTTTCGTTCCACCGTATTTCAAAATACCGATGAAGCGGCATGTCMNT 1647
      ||| ::::::::::::::::::::|
487 InSerPheAsnLysIleAlaAlaThrAspSerGlyAlaValIle..... 501
1648 CATATATGCCACACATCCACCGTTACCATTTACAGGGAATGAAGTAT 1697
      ||| ::::::::::::::::::::|
502 .....ThrsnThrSerThr..... 506
1698 TACACACACGAGTGTAGACATATTCATAGACTTAATACAGCAAGAAA 1747
      ||| ::::::::::::::::::::|
507 .....LysLysSerIleLeuSerLeuAsnThrAlaAspT 519
1748 TTGCTTACACAGGTGTGGTTGGCGAAGAAATACGACCAAAACGAGG 1797
      ||| ::::::::::::::::::::|
519 yrlLysThrIleGly.....AsnIleAsnGly 527
1798 CGGCTACACCTGTT.....TACACGCCGCGCGAGAACCGCACCN 1841
      ||| ::::::::::::::::::::|
528 AsnLeuAspValLeuGlnHisGlnThrLysGlnAsnArgArgIle 544
1842 GCGGCTTCCGCGGCAACAAATTTAAACGCGACATCAGCGCAACAAACG 1891
      ||| ::::::::::::::::::::|
544 uileLeuAspGlyGlyValAspThrThrAsnAspIleSerLeuAspAsnT 561
1892 GCAAACTGTTTTCAGCGCGACGCGACACCGCGCTACAAATCATTTA 1941
      ||| ::::::::::::::::::::|
561 hGlnLeuSerMetGlnGlnHisAlaThrGlnHisAlaIleLysArgAsp 577
1942 GGAAGC.....GGGTC 1952
      ||| ::::::::::::::::::::|
578 GlyAlaPheSerCysSerLeuProAlaPrometArgPheLeuGlySe 594
1953 GTCAAAAATGGAAGTATC.....CCACAGGAGAAATCGTGTGGACA 1996
      ::::::::::::::::::::|
594 rAspArgValAlaGlnGlyMetGlnAsnThrGlnAlaAspAlaValLysGln 611
1997 AC..... 1998
      |||
611 snGlyAsnAlaLysLysThrAsnAsnAlaValSerAspLeuSerGlnPro 627
1999 GACTGGATCNCGCCGACGTTTAAACGGAATTTCCATTCAGGCGG 2048
      ||| ::::::::::::::::::::|
628 AspArgGlnThrGlyLysThrPheArgPheGlyLysLeuHisLeuGlnAsnSe 644
2049 GCAGCGGTGATTTCCGCAATGTGCAAGTGAAGGAGCGATGNCMT 2098
      ::::::::::::::::::::|
644 rAspPheSerValGlnArgAsn..AlaAsnValIleGlyAspIleGlnA 660
2099 TGAGCAATCAGGCCCAAGCACTT..... 2121
      ||| ::::::::::::::::::::|
660 lAserLysSerAsnIleThrIleGlyAspThrThrAlaLysIleAspLeu 676
2122 .....TTGGTGTGCGACCGCAT.. 2139
      ||| ::::::::::::::::::::|
677 HisAlaGlyLysAsnIleThrGlyAspGlyPheArgGlnAsnIle 693
2140 .....CAAGGCATACCAATCTGTACACGTTGCGATCGGA 2173

```

```

693 eValArgGlyAsnSerGlnGlyLysThrLeu.....PheT 705
      ||| ::::::::::::::::::::|
2174 CNGGCTGTGCAAAATGTGTGCAANAANAACATTACGACAGATTAAGTATT 2223
      ||| ::::::::::::::::::::|
705 hGlyGlyLysThrAlaGlnAspSerThrIle.....ValIle 717
2224 GCTTCATTTGATAGACAGCAGTTCAGGCAATGTCAGTTCAGTTCAGTTC 2273
      ::::::::::::::::::::|
718 LysAspArgLysAlaLysAlaLeuPheSerAsnArgValLysLeu..... 732
2274 CGTNTNTTNAANCTCNCNGGCGTGCNNCACTNAANGCAATCTTAGTG 2323
      ::::::::::::::::::::|
732 ..... 732
2324 CAATGCGCATATACAGTATATAGTACAGCCACAAAGCC.....ACC 2364
      ::::::::::::::::::::|
733 .....AsnThrLysAlaThrIleGlnAsnGlyAlaAspValThrThr 746
2365 CAAAAGCGCAACCTTAGCCTGTGGGCAATGCCCAAGCAACATTTAATCA 2414
      ||| ::::::::::::::::::::|
747 GlnSerGlyMetPheSerThr.....SerAs 755
2415 AGCCCATTTAAAGCGCAACNCATCGGNTTCGGGCAAT..... 2451
      ::::::::::::::::::::|
755 pIleSerIleSerGlyAsnLeuSerMetThrGlyAsnProAspLysAspA 772
2452 .....GCTTCATTTAAATCTAAGC 2469
      ||| ::::::::::::::::::::|
772 snLysPheGlnProSerIleThrLeuAsnAspAlaSerThrLysLeuThr 788
2470 AACAAGCGCGGCAACAAGCGCATGTACGCTTTCCGCAACGCTTAAGC 2519
      ::::::::::::::::::::|
789 AspAspSerAlaArg.....LeuValAlaLysAsnLysAl 800
2520 AAAGCTA.....AGCATTCGCGACATCAAGCGCAATGTCGCGCTAG 2560
      ||| ::::::::::::::::::::|
800 aserValAlaGlnLysPheIleHisSerThrLysSerAlaSerIleThrPheG 817
2561 CCGATTAAGCGCATTTCCATTTGAAACAGCCGCTTACCGCAACATC 2610
      ||| ::::::::::::::::::::|
817 Lys.....HisAspLysSerAsp.....LeuSerGlnLeu 826
2611 AGCGGC.....AGCAAGANACGCAATTAACATTAACAGCAGCATG 2654
      ||| ::::::::::::::::::::|
827 SerAspArgThrSerLysGlyLeuAlaLeuGlyLeu..... 839
2655 GACGCTCGCTCAGGACGCAAGATTA.....GCAATTTAAACCTTG 2695
      ||| ::::::::::::::::::::|
840 .....GlyLysPheAspValSerTyrArgGlySerValAsnAlaP 853
2696 ACAAGCGCACCATTTACATCAATTCGCGCTATGCGCAATGCTGCGAGC 2745
      ::::::::::::::::::::|
853 roSerAlaSerAlaThrMetAsnAsnThrTrp..... 863
2746 GCGCAACCGCGAGNGTGTACAGACGCGCGCGCGCGCTTCGCGCGGTTG 2795
      ||| ::::::::::::::::::::|
864 TrpIleLeuThrGlnLysAspSerAlaLeuLysThrLeuLysSerThrAsnSe 880
2796 CTAATTTCCGTTACACCGCGCAACTTCGAGAAATCCGTTTACACAGC 2845
      ::::::::::::::::::::|
880 rMetValLysThrPheThrAspSerAlaAsn..AsnLysLysPheHisThrL 896
2846 TGACGGTGAACGGCAATTTGAACNGTCAGAGACATTCGCGCTTAAAGTGG 2895
      ||| ::::::::::::::::::::|
896 euthrValAspGlyLeuAlaThrSerAsnSerAlaLysThrAlaLysThr 912
2896 GAACTCTTCGCTACCGAAGCAACAAATTTGAAGCTGCGGAAAGTTCCGA 2945
      ::::::::::::::::::::|
913 AsnLeu.....serGlnSerAspLysLeuGlnVal..... 922
2946 AGNNACTTACACTTGGCGGTCAACAATTCACGCGCAAGCAACCGCTAAGCC 2995

```

```

922 ..... 922
2996 TCATCAATGACGTAGTGAAGCAAGCAACAAACCGCTGCCAA 3045
      : : : : :
923 ..... LysIysHisLeuSerGly 928
3046 AACCTTAATTCACCTGCAGAAACGACGTGATGCCGGCGGTGGC 3095
      : : : : :
929 GLuSnSnIleLeu..... ValAspPhe..... 937
3096 TTACCACTATCCGCAAGACGGCGAGTCCGCTCATATACCGGTCA 3145
      : : : : :
938 ..... LeuGlnIysProThr 943
3146 AAGAACAAGACCTTCGACAACTCGCAAGCAGAGCAAAAAACAG 3195
      : : : : :
943 roGluIysGlnLeuAsnIleGlu..... 950
3196 GCGGAAAAAGCAACGGCGCAAGCCTTGACGCGCTGATGGCGCGGCG 3245
      : : : : :
951 ..... LeuValSerAlaProly 956
3246 CGATGCCCGCCAAAAGACAGAAACGCTTCCGGAACCGCCGCGNGCGAG 3295
      : : : : :
956 SASPThrSnGlnIysnValIlePheIys..... AlaS 966
3296 GCGGGGAAATGTGCGCATTTATGACGGCGGAGAGAAAGAAAAAGGGTG 3345
      : : : : :
966 eIysGlnThrIleGlyPhe..... 972
3346 CAGCGGATAAAGACAGCGCNTTGGCGAAGACGCGAGCGAAACCGC 3395
      : : : : :
973 ..... SerAspValThr 976
3396 GCCGNTACACCGCGCTTCCCCCGCGCGCGCGCGGGATTTGC 3445
      : : : : :
976 rProValIleThr..... 981
3446 CGCAACCGCAGCCCAACCGCACTCAACCCCAACCGCAGCGAGCTG 3495
      : : : : :
982 ..... ArgGlnThr 984
3496 ATNACCGCTTATGCCAATAGCGGTTTGATTTTCCGCGAGCTCAA 3545
      : : : : :
985 AspAspIysIleThrTrpSer... LeuThrGlyTrpAsnThrValAlaAs 1000
3546 CAGCGTTTTCGCGCTACAGAGCAATTTGACCGCGTGTTCGCCGAAGAC 3595
      : : : : :
1000 nIys..... GluAlaTr 1004
3596 GCCGACGCGCTTGGACAGACGATCCGCGNACGCAACACATACCGT 3645
      : : : : :
1004 hrArgAsnIleAlaIleLeuPheSerVal..... 1013
3646 TCGCAAGATTTCCGCGCTAC..... CGCA 3671
      : : : : :
1014 ..... AspTyrIysAlaPheLeuAsnGlnValAsnAsnLeuAsnIlyAsr 1028
3672 ACAAAACCGCCTGGCCAAATCGTATGCAAAAAACCTC..... 3711
1028 gmetGlyAspLeuArgAspIleAsnGlnGluAlaGlyAlaTrpAlaArgI 1045
3712 ..... GCGAGCGCGCGCTGGCGATCTGTTTTCGCAACGCGGAC 3753
      : : : : :
1045 IeMetSerGlyThrGlySerAlaSerGlyGlyPheSerAspAsnTyrThr 1061
3754 GAAACACNTTGCAGCAGCGCATCGCAACTCGCAGCGCTTGCACGCG 3803
      : : : : :
1062 His..... ValGlnValGlyValAspIysHisGlnLeuAspGlyLe 1076
3804 CCGCGTTTTCGCGCATACGCGATCGCGAGGTTGCGCATCGGCATC 3849
      : : : : :
1076 unspleupherThrGlyPheThrValThrHisThrAspSerSerAlaSerA 1093

```

```

3850 ..... AGCAGCGCGCGGGT..... 3864
1093 IAspValPheSerGlyIysThrIysSerValGlyAlaGlyLeuThrAla 1109
3865 ..... TTACGACGCGCANTCTNTCAGACGATCGAGGCAAAAT 3905
      : : : : :
1110 SerAlaMetPheAspSerGlyAlaTrpIleAspLeuIleGly..... 1123
3906 CCGCGCGCGCGTGTGCATTTACGCGATTCAGGACAGATACCGCGCGT 3955
      : : : : :
1124 ..... LysTyrValHisHis..... AspAsnGlnTyrThrAlaTrp 1136
3956 TCGCGCGATTCGCGCATCGAACCGTACATTCGCGCAACGCGCATTTGTC 4005
      : : : : :
1136 heAlaGlyLeuGlyThrArgAspTyr..... 1144
4006 CAAAAAGCGATTAACCGTACGAAAAACGTCAATATCGCACCCCGGTCT 4055
1144 ..... 1144
4056 TCGCTTCAACCGNTACGCGCGCATTAAGCAGCATTAATTCATTCAAAC 4105
1145 .SerThrHisSerTrpTyrAlaGlyAlaGlnIleGlyTyrArgTyrHisV 1161
4106 CCGCGCAACACATNTCCATCACNCCTTATTNAGCGCTGCTATACCGAT 4155
      : : : : :
1161 alThrGlnAspAlaTrpIleGluProGlnAlaGlnLeuValIlyr... Gly 1176
4156 GCCGTTTCGCGCAACGTC..... 4173
1177 SerValSerGlyIysGlnIleAlaTrpIlyAspGlnGlyMetHisLeuSe 1193
4174 ..... CGAACACCGCTCAATA 4189
1193 rMetIysAspIysAspTyrAsnProLeuIleGlyArgThrGlnVal.... 1208
4190 CCGCGNTATTTGCTCAGGATTTTCGCAAAACCGCGAGTCG... GAATGG 4236
1209 ..... AspValGlyIysSerPheSerGlyIysAspTrp 1219
4237 GCGTAAACGCC 4248
1220 LysValThrAla 1223
seq_name: p1r2:B41500

```

```

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 2 - Haemophilus influenzae (fra
N:Alternate names: immunoglobulin A1 proteinase type 2
C:Species: Haemophilus influenzae
C:Date: 30-Jun-1992 #sequence-revision 30-Jun-1992 #text-change 08-Dec-2000
C:Accession: B41500
R:Grundy, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A:Title: Localization of the cleavage site specificity determinant of Haemophilus inf
A:Reference number: M41500; MUID:90129281
A:Accession: B41500
A:status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-384 <GRU>
C:Comment: This Iga1 proteinase is classified as type 2 because it cleaves at a proli
C:superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase, metalloproteinase

```

```

alignment_scores:
Quality: 444.50 Length: 426
Ratio: 1.837 Gaps: 22
Percent Similarity: 56.808 Percent Identity: 32.160

```

```

alignment_block:
us-09-303-518d-651 x B41500 ..

```

Align seg 1/1 to: B41500 from: 1 to: 384

```

64 GGGTTCGGCTGCTACTTACCATATGCGCTGCTGGCATTCCTTCC 113
   : : : : : : : : : : : : : : : : : : : : : : : :
5 LysPheLysLeuAsnPhelLeuAlaLeuThrValAlaLysAlaLeuThrPr 21
   : : : : : : : : : : : : : : : : : : : : : : : :
114 CCAAGCTTGGCGGACACACTTATTTCGGCATCATCACTCACTACTATC 163
   : : : : : : : : : : : : : : : : : : : : : : : :
21 oTyThrGluAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePheA 38
   : : : : : : : : : : : : : : : : : : : : : : : :
164 GCGACTTGGCGGAAATTAAGCAAGTTGCGACTCGGGCGGAAGATATT 213
   : : : : : : : : : : : : : : : : : : : : : : : :
38 rGAspPheAlaGluAsnLysGlyAlaArgPheSerValGlyAlaThrAsnVal 54
   : : : : : : : : : : : : : : : : : : : : : : : :
214 GAGTGTACACAAACAAAGGGAGTGGTCGGCAATCATGACAAAC 263
   : : : : : : : : : : : : : : : : : : : : : : : :
55 GluValAlaArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
   : : : : : : : : : : : : : : : : : : : : : : : :
264 C...CGAGTATGATTTTCTGTGTGTGCGGTAC...GGCGTGGCG 307
   : : : : : : : : : : : : : : : : : : : : : : : :
71 yIleProMetIleAspPheSerValAlaAspValAsnLysAlaGlyT 88
   : : : : : : : : : : : : : : : : : : : : : : : :
308 CATGTGGGCGCATCAATATATTGTGAGGTGCGCATACGGCGCTAT 357
   : : : : : : : : : : : : : : : : : : : : : : : :
88 hIleuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104
   : : : : : : : : : : : : : : : : : : : : : : : :
358 AACACGTTGATTTT...GGTCGGAGAGAAAGNA 389
   : : : : : : : : : : : : : : : : : : : : : : : :
105 MetAsn...AspPheTyrPheGlyHisTyrAsnGlyHisAlaGAspValSe 120
   : : : : : : : : : : : : : : : : : : : : : : : :
390 TCCCGATCAGCACCGCTTTTCTTACCAATGTGAAGAAATATTTATA 439
   : : : : : : : : : : : : : : : : : : : : : : : :
120 rAspAspGluAsnLys...TyrSerValValThrGlnAsnAsnValAla 135
   : : : : : : : : : : : : : : : : : : : : : : : :
440 AGCCT...GACAATTCACACCTTACAC...GGCGATTAACATATG 480
   : : : : : : : : : : : : : : : : : : : : : : : :
135 snProAsnGluAsnTrpHisValAspLysAlaGluAspAspTyrAsnMet 151
   : : : : : : : : : : : : : : : : : : : : : : : :
481 CCGCGTTGCATTAATTTGTGCACAGATGACAGACTGTGCAATGCAG 530
   : : : : : : : : : : : : : : : : : : : : : : : :
152 ProArgLeuAsnLysPheValThrGluValAlaProThr...ThrPr 166
   : : : : : : : : : : : : : : : : : : : : : : : :
531 TGCATGAGAGGGGAAAT...ACCTATTCGATTAAGAAATATATC 571
   : : : : : : : : : : : : : : : : : : : : : : : :
166 oThrLeuAlaGlyAspAspLeuGlnThrTyrLysAspLysGlnLysTyrL 183
   : : : : : : : : : : : : : : : : : : : : : : : :
572 CCGAGCGTCCGATCGGCTCAGA...GGCGATTAACATATG 597
   : : : : : : : : : : : : : : : : : : : : : : : :
183 euSerPheValAlaValGlyAlaGlyAlaGlnLeuValTyrGlnLysGly 199
   : : : : : : : : : : : : : : : : : : : : : : : :
598 ...CACCACTATTTGGCGTTATGATGATGACAAACAGCGC... 633
   : : : : : : : : : : : : : : : : : : : : : : : :
200 SerHisHis...ValGluAspLysGlnHisGlyGluAspLeuLys 213
   : : : : : : : : : : : : : : : : : : : : : : : :
634 GATTATCTACTACCGCGCATGTTAATGGCGGCAATACAT... 678
   : : : : : : : : : : : : : : : : : : : : : : : :
213 sasPheSerAlaAlaTyrAlaTyrAlaIleGlyIleThrProTyrLysG 230
   : : : : : : : : : : : : : : : : : : : : : : : :
679 ...ATGACAGGTTGGGA 693
   : : : : : : : : : : : : : : : : : : : : : : : :
230 yIleAsnIleAspProSerGlnSerLysGlyLeuIleGlyPheGly 246
   : : : : : : : : : : : : : : : : : : : : : : : :
694 AAT...AATGGCTAATTAATGTTGAGCGGCA 722
   : : : : : : : : : : : : : : : : : : : : : : : :
247 AspSerAlaGluAspHisValIleAsnSerLysThrLeuLeuSerGlnAs 263
   : : : : : : : : : : : : : : : : : : : : : : : :
723 TGTGCGCATGCCAAGCACTATGGCCCTATGCGGATTCGAGTCCGCGAG 772
   : : : : : : : : : : : : : : : : : : : : : : : :
263 P...ProLeuThrAsnTyrGlyValLeu 272
   : : : : : : : : : : : : : : : : : : : : : : : :
773 GCGACAGCGGTGCGCAATGTTATTATGACAAACAAACATTAATG 822
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

272 LysAspSerGlySerProLeuPheAlaPheAspLysGlnGlnAsnLysTyr 288
823 CTGCTCAACGAGGATTTTACAAACCGGCTACCTTATTCGCGACGGA 872
   : : : : : : : : : : : : : : : : : : : : : : : :
289 PhePheIleGlyProTyrThrTyrTrpAlaGlyTyrGlyLysLysSerHis 305
   : : : : : : : : : : : : : : : : : : : : : : : :
873 CGGTTTCACAGCTGATACGCAAGATTTGTTCTGATGATCATTTACAG 922
   : : : : : : : : : : : : : : : : : : : : : : : :
305 rGlnGlnIleTrpAsnIleTyrLysSerGlnPheThrLysAspValLeuAsn 322
   : : : : : : : : : : : : : : : : : : : : : : : :
923 GCGATACACATACCGCTCTNTTTGAACCGGCGAGTACAGGACAT... 966
   : : : : : : : : : : : : : : : : : : : : : : : :
322 yAspSerAlaGlyLeuLeu...LysGlyHisThrGln 333
   : : : : : : : : : : : : : : : : : : : : : : : :
967 TTTTCCCTTATCATCCAAACAAACAGTACGGGTACGATACGAAACCA 1016
   : : : : : : : : : : : : : : : : : : : : : : : :
334 TyrAsnTrpPheSerHisGlnLysAsn...ThrSerMetIleSerAsnGlySe 349
   : : : : : : : : : : : : : : : : : : : : : : : :
1017 CGAAAGGTTTCCATTCACAAAGCTTAAGTACAGACAGTCCGACTGTTG 1066
   : : : : : : : : : : : : : : : : : : : : : : : :
349 rGluLeuLeuGly...ValAsnLeuPheA 358
   : : : : : : : : : : : : : : : : : : : : : : : :
1067 ACCAATCTTTGAATGAAGTAAAGTAAAGACAGTTCAGCGGCGAGGCT 1116
   : : : : : : : : : : : : : : : : : : : : : : : :
358 spAsnSer...LysHisThrAsnArgGlu... 366
   : : : : : : : : : : : : : : : : : : : : : : : :
1117 GTTAATCAGTACCGCTTCAAGGTTAAACAAACGSGTGAACCTTCTTTAT 1166
   : : : : : : : : : : : : : : : : : : : : : : : :
367 ...GlnGlyAsnGlyThrLeuThrLeu 384
   : : : : : : : : : : : : : : : : : : : : : : : :
1167 CGATTACGCGACGCGCAACCTCATCTTA 1194
377 ...GlnGlyAsnGlyThrLeuThrLeu 384
seq.name: p1r2.A65044
seq documentation block:
hypoetical protein b2647 - Escherichia coli (strain K-12)
C.Species: Escherichia coli
C.Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C.Accession: A65044
R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Riley, M.
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A.Title: The complete genome sequence of Escherichia coli K-12.
A.Reference number: A64720; M01D:9742617
A.Accession: A65044
A.Status: preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-1569 <BIAT>
A.Cross-references: GB:AE000350; GB:U00096; NID:92367147; PIDN:AMC75695.1; PID:g17896
A.Experimental source: strain K-12, substrain Mgi655
alignment_scores:
Quality: 372.50 Length: 1676
Ratio: 0.491 Gaps: 76
Percent Similarity: 45.286 Percent Identity: 19.153
alignment_block:
US-09-303-518D-651 x A65044 ..
Align seg 1/1 to: A65044 from: 1 to: 1569
127 GCACACACTTATTTGGCATCACTACCAATACTATCGGACTTGGCGA 176
   : : : : : : : : : : : : : : : : : : : : : : : :
160 GlyThrThrIleGluSerGlyAsnGlnAspValTyrLysGlyGlyIleSe 176
   : : : : : : : : : : : : : : : : : : : : : : : :
177 AAATTAAGGCAAGTTTGCAGTCCGGCGGCAAGATATGAGGTTTACACA 226
   : : : : : : : : : : : : : : : : : : : : : : : :
176 rAsnGlyThrThrIleLysGlyGlyAlaSerAlaGlyAlaGlyGlySerA 193
   : : : : : : : : : : : : : : : : : : : : : : : :
227 AAAAAGGGAGTTGGTC...GGCAATCAATGACAAACGCCCG... 267

```



```

1828 GAGAGCGCAGCCCGCTTCCGGGAGACAATTTAAACGGCAACAT 1877
    : : : : : : : : : : : : : : : : : : : : : : : :
672 SerAspLysAlaSerIleLysGlyLysGlnThrValTyrLysLeuAl 688
1878 CAGCAACAACAAC.....GGCAACGCTTTTTCAGCGGAGACCGA 1918
    : : : : : : : : : : : : : : : : : : : : : : : :
688 aThrGluAlaSerIleGlnSerIleGlnIleValAspGlyLysSerI 705
1919 CACCGCAGCGCTTACATTCATTAGAACGGCGTGTCAAAAAGGAAAGT 1968
    : : : : : : : : : : : : : : : : : : : : : : : :
705 hrcGluLys.....ThnHisIleAsnGlyLThrGlnThrValGlnAsn 719
1969 ATCCCAACAGGA.....GAATCGTGTGGACACAGCACTGATCA 2009
    : : : : : : : : : : : : : : : : : : : : : : : :
720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnGlnIle 736
2010 CCGCAGCTTTAAAGCGAAATTCATTCATTCAGCGCGGAGCGGTGA 2059
    : : : : : : : : : : : : : : : : : : : : : : : :
736 tAlaAsnGlyThrAlaGlySerIleIleAsnGlyLysSerGlnVal 753
2060 TTTCCCGC..... 2067
753 aLAsnGlyLysGlyLeuAlaGlnAsnSerValLeuAsnAspGlyLThr 769
2068 ...AATGTCGCAAGTGAAGCGCATTCGATTCGAGCAATCAGCCCA 2114
    : : : : : : : : : : : : : : : : : : : : : : : :
770 LeuAspValAlaArgLysGlySerAlaThrGlyIleGlnGlnSerSer 786
2115 ASCAGTTTGGTGTGCACGCGATCAACGCAATCACTTACACGTT 2164
    : : : : : : : : : : : : : : : : : : : : : : : :
786 nGlyAlaLeuValAlaThrThrAlaThrAlaThrGlyThrAlaArg 803
2165 CGAGCTGACN.....GCTCTGACAAAT..... 2187
    : : : : : : : : : : : : : : : : : : : : : : : :
803 lAspGlyAlaAlaPheSerIleGlnGlnAlaAlaAsnIleLeu 819
2188 .....TGTCGCAANAANCAATTACCGAGCATPA 2216
    : : : : : : : : : : : : : : : : : : : : : : : :
820 LeuAlaAsnGlyLysValLeuThrValGlnSerAspThrSerSerAsp 836
2217 AGTGATGCTTCATG.....ACTAAGANGACN 2245
    : : : : : : : : : : : : : : : : : : : : : : : :
836 sThrGlnValAsnMetGlyLysArgGlnIleValLysThrLysAlaThr 853
2246 TNAAGCGCANGTNAGCTNNCCNATNACGNTNNTNAANCNCNGG 2295
    : : : : : : : : : : : : : : : : : : : : : : : :
853 lAthrGly.....ThrThrLeuThrGly 861
2296 CNTGCNCACTNAANGC..... 2313
    : : : : : : : : : : : : : : : : : : : : : : : :
862 GluGlnIleValGlnGlyValAlaAsnGlnThrThrIleAsnAspGly 878
2314 .....AATCTTAGTGCAAATGGCGAT.....ACAGTTATACAG 2347
    : : : : : : : : : : : : : : : : : : : : : : : :
878 yIleGlnThrValSerIleAsnGlyLysAlaIleLysThrLysIleAsn 895
2348 TCAGCCACAAGCCACCCAAAACGAGCAAC.....CTTAGCCTGTG 2388
    : : : : : : : : : : : : : : : : : : : : : : : :
895 lueGlyLysThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleVal 911
2389 GGCATATGCCAAGCAACATTTAATCAAGCCATTAAGCGCAACNCATC 2438
    : : : : : : : : : : : : : : : : : : : : : : : :
912 GlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly...IleG 927
2439 GGNTTGGGCAAT.....GCTTCATTTAATCTAAGCAACAACGCG 2479
    : : : : : : : : : : : : : : : : : : : : : : : :
927 uIleSerGlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeu 944
2480 CACAACAAC.....GGCAGTGTGAGCGCTTCGAGCAAC 2511
    : : : : : : : : : : : : : : : : : : : : : : : :
944 lAthrAsnMetLeuLeuGlnLysAsnGlyLysAsnLeuValLeuAlaGly 960

```

```

2512 GCTAAGCGCAAGCTAAGCACTTCCGCACTCAACGCAATGTCTCCTAGC 2561
    : : : : : : : : : : : : : : : : : : : : : : : :
961 ThrGluAlaArgAspSerThrValGlyLysGly..... 972
2562 CGATAGGCGATATTCATTTTGAANAACAGCGGCTTACCGAGACATCA 2611
    : : : : : : : : : : : : : : : : : : : : : : : :
973 .....AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsn 987
2612 GCGGAGCAGAGGAAACAGCATTCACCTTAAAGACGCAATGAGCGGTG 2661
    : : : : : : : : : : : : : : : : : : : : : : : :
987 erGlyGlyLysThrLeuGlyArgSerLysAspGluPheGlnAlaLeu 1003
2662 CCGTAGCGCAGCAATTA..... 2679
    : : : : : : : : : : : : : : : : : : : : : : : :
1004 AlaArgAlaGluAspLeuGlnValAlaGlyLysThrAlaIleValTyrAl 1020
2680 .....GGCAATTTAAAC 2692
1020 aGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSer 1037
2693 TT..... 2694
1037 eumetThrProArgAspAsnValThrProValLysLeuGlnGlyAlaVal 1053
2695 .....GACAAGCCACCATTAACATTCACCTT..... 2727
1054 ArgIleThrAspSerAlaThrLeuThrLeuLysAsnGlyAlaAspThr 1070
2728 ..CGCCAGCATCTGCAGCGCGCAACCGCAGCGT..... 2763
    : : : : : : : : : : : : : : : : : : : : : : : :
1070 rLeuAlaAspLeuThrAlaIleSerArgGlySerValThrLeuAsnSer 1087
2764 .....TCAGACAGCGCGCGCGCGCTTCCGCTT 2799
    : : : : : : : : : : : : : : : : : : : : : : : :
1087 snAsnSerCysAlaGlyThrSerAsnGlyLysThrValAsnSerLeu 1103
2800 TTA.....TCGCTTACCGCCCAAC 2819
    : : : : : : : : : : : : : : : : : : : : : : : :
1104 LeuLeuAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaAlaProAl 1120
2820 TTCGCTGAATCCCGTTTCAACAGCGTGAAGCAAAATGTAACN 2869
    : : : : : : : : : : : : : : : : : : : : : : : :
1120 aThrThrAsnGlyLysThrAsnThrLeuThrAsn...GlnLeuSer 1136
2870 GTCAGGACATTCGCTTATGTGCGAAGCTTTCGCTACCGAAGCAG 2919
    : : : : : : : : : : : : : : : : : : : : : : : :
1136 LysGlyAsnPheTyrLeuHisThrAsnValAlaGlySerAlaGlyAsp 1152
2920 AATTTGAAGCTGGCGGAAGTCCGAAGNACTTACACCTTGCGGTCAA 2969
    : : : : : : : : : : : : : : : : : : : : : : : :
1153 GlnLeuValAlaAsnAsnAlaThrGlnLysAsnPheLysIlePheValG 1169
2970 CAATACCGGCAAGAACCCCTAAGCCTCGATCAATTGACGGTAGTGAA 3019
    : : : : : : : : : : : : : : : : : : : : : : : :
1169 nAspThrGlyValSerProGlnSerAspAlaMetThrLeuVal.... 1184
3020 GGAAGACAGCAACAGCGCTGTCCGAACCTTAATTCACCCGCAAAAC 3069
    : : : : : : : : : : : : : : : : : : : : : : : :
1185 .....LysThrGlyGlyLysAlaSerPheThrLeuGlyAsn 1197
3070 GAA.....CAGGTGATCGCGGCGGTGAGCTTACCAATCGACGCA 3113
    : : : : : : : : : : : : : : : : : : : : : : : :
1198 ThrGlyLysPheValAspLeuGlyThrTyrGlyLys...ValLeuLys 1213
3114 AGACGGC.....GAGTCCGCTGATATCCGCTCAAGAACAGAC 3157
    : : : : : : : : : : : : : : : : : : : : : : : :
1213 lAspGlyAsnSerAsnThrPheLeuThrAsnAspValLys..... 1226
3158 TTTCGACAACACTGGCAAGGAGCAAGCAAAACAGCGCGAAGAAC 3207
    : : : : : : : : : : : : : : : : : : : : : : : :
1226 ..... 1226
3208 AACGGCAAGCCTTGACCGCTGATGGCGCGCGGAGATGCCGCCGA 3257

```



```
289 TCCTGTAACGGCGGTGGCGCATTTGATG..... 315
217 AsnlyserglyserglinspvalvalglnlyserleuAlaThrAsnTh 233
316 .....GGCGATCAATATATAT.....GTGA 334
233 rThrIleAsnlyglYArgglInlyValgluInserThrValgluThrT 250
335 GCGTGACATACGGCGGCTATACACGCTT..... 366
250 hrThrIlelyAsnlyglYgluInlyrValylrgluSerArgAlaLeu 266
367 .....GATTTGGTGGGAGAGACNAATCCGATGACGACCG 404
267 AspThrThrIlegluInlyglYThrInlyserleuAsnlyserThrAl 283
405 TTTTCTTACCAATTTGTGAAGAATAATTAAGCTGACATATTCAC 454
283 elysAsnThrInlyserIserIyglYThrInlyleleAspAsnThr. 299
455 ACCCTTACAACGGCGATTCATATATGCCGTTTGATTAATTTGCACA 504
300 .....Ser 300
505 GATGCAACCTGTGCAATGACAGT.....GACATGAGGG 542
301 SerSerAspValIlegluValIlyserIyglYValleuAspValserI 317
543 GAATACCTTTCGATTAAGAAAATATCCAGCGTGTCCGATCGGCT 592
317 yglYThrAlaThrAsnValThrIn..... 325
593 CAGGACACCATATTGGCGTTATGATGATGACAAACGCGGATTTATCC 642
325 ..... 325
643 TACTCGCGGCATGTTAATGGCGCAATACAT.....ATGCA 683
326 HisAspGlyAla...IleleuLysThrAsnThrAsnlyThrThrValSe 341
684 GGGTTGGGGAATTAATGGCTANTTAGTTG..... 714
341 rGlyThrAsnsergluInlyAlaIleHisAsnHisValAlaAspA 358
715 .....ACGCGCATGTGCGCATGCCCACGACTATGCGCCT 750
358 snValleuInlyAsnlyglYHisleuAspIleAsnAlaIyglY... 373
751 ATGCCGATTGCAGTGGCGGACGCGACAGCGGTTGCCAATGTTATTGA 800
374 .....serAlaAsnLysThrIleIlely 381
801 TGACAAAACAATTAATGCTGCTCAACGAGCTTTACAAACCGGCT 850
381 sAspLysglYThrMetSerValleuThrAsnAlaIyAlaAspAla... 396
851 ACCCTTATCCGGGACGAAACGTTTCCAGCTGATACGCAAGATTTGG 900
397 .....ThrArgIleAspAsnlyglYValMet..... 405
901 TTCTACGATGACATTTACAGAGCGATACATACCGCTTTTGTGAAC 950
406 .....AspValAlaIyAlaThrAsnThrIleIle..... 416
951 GCGCAGTACGACATTTTCTTACATCCACACGCTACGCGGTA 1000
417 .....AsnlyglY...ThrGlnAsnIleAsnAsnlyrGlyleuAlat 430
1001 CGGTAAACAACAACGAAAGTNTCCATCCAAAGCTT.....AAA 1044
430 hrGlyThrAsnIleAsnserIyThrInlyleLysSerIyglYlys 446

1045 GTACAGACATCCGACTGTGTAACAATCTTGAATGAACGATTAAGA 1094
447 AlaAspThrThrIleIleSerSerIySerArgInlyValgluLysAs 463
1095 ACCAGT.....TACGGCAGGGGTGTATATCAGTACC 1129
463 polYThrAlaIleglYserAsnIleSerAlaIy..... 475
1130 GTCAAGGTTAAACAAGGTGAAAACCTTCTTATCATGATTAACGCAAC 1179
476 .....SerleuIleValIyThrIy 482
1180 GCGAACATCATCTTATCAACACATCAACAGGCGGCGGCTTTGTA 1229
483 Gly.....IleAlaHisglYAlaAsnInlyThrIySerAlaIe 496
1230 TTTGAAGTGATTTTACGGTCTCGCCCGAACAACAACAGTGGCAAG 1279
496 uValAlaAsn.....ThrG 501
1280 GCGCGGCGCTTCATATCAGT.....GAAGACGATACCGTT 1314
501 IyAlaIyThrAspIlegluInlyrAsnLysleuSerHisPheThrIle 517
1315 ACTTGAAGAATTAACGGCGGTGCAACAGCGCCTGTCCAAATGGCAA 1364
518 Thr.....GlyglYAlaAlaAsnlyrValValleuInly...AsnTh 530
1365 AGGACGCTGCAGCTTCAGACCAAGGGGAAACCAAGGCTCATGACGG 1414
530 rGlyleuThrValValAlaLysThrSerAlaIyAsnThrThrIleA 547
1415 TGGGCGAGGTACAGTATTGATGATCAGCAGCAGATTAAGGCAAA 1464
547 spAlaIyglYLysleuIleValglInlysgluAlaLys..... 559
1465 AAACAAGCCTTATGTAATCGCTGNTCAAGCGGCGGATACGGTCA 1514
560 .....ThrAspSerThrAr 564
1515 ACTGATATCCGATTAATCAGTTCAACCCGACAAACCTATTTGGCTTC 1564
564 glLeuAsn.....A 567
1565 GCGGCGCGCTTTGATTTAAAGGCGATTCGTTTCACACCGTATT 1614
567 snGlyglYValleuInlyValglInlyAspIyglYAlaLysHis...Val 582
1615 CAAATACCGATGAAGGCGGATGATGNCNATCATTAATGCCACAAAC 1664
583 gluInlyInserIyglYAlaIleuIle.....AlaSerThrTh 595
1665 ATCCACCGTTACATTACAGGGAATGAAGT.....A 1696
595 rSerGlyThrleuIlegluInlyThrAsnserIyrgIyAspAlaPheTyI 612
1697 TTACACAACGAGTGTAAAGATATATCAATGATTAATTAACACCAAGA 1746
612 leArgAsnsergluAlaIyAsnValValleuInlysnAlaIySerleu 628
1747 ATTGCTTACACAGGTTGGTGGCGAGAAAGATACGACCAAAACGACGG 1796
629 ThrValValThrIySerArgAlaValAspThrIleIleAsnAlaAsnI 645
1797 GCGGCTCAACCTT..... 1809
645 yLysMetAspValTyrglyLysAspValglYThrValleuAsnSerAlaG 662
1810 .....GTTTACAGCCGCGCGACAGACGACCCNGCTGCTT 1848
662 IyThrGlnThrIleTyrrAlaSerAlaThrSerAspLysAlaAsnIleLys 678
1849 TCCTGGCGAACAATTTAAACGGCAACATCACGCAAAACAAAC..... 1890
```



```

3303 AAATGCGCATTTATCGAGCGGAGAGAAAAAGGGTGCAGCGG 3352
1234 ..... 1234
3353 ATAAAGACAGCGCTTTGGCGAAGCGCGAAGCGGAAACCGCGCGGNT 3402
1235 .....ProAsnProAs 1238
3403 ACCAGCCGCTTCCCGCGCGCGCGCGCGCGGATTTGGCGGACAC 3452
1238 nProLysProAsp..ProLysProAspProLysPro..Asp...ProLysPr 1253
3453 GCAAGCCCAACCGCAACCTCAACCCGACGCGACCTGATNAGCC 3502
1253 oAspProThrProGluProThrProThrProValProGluLysArgLeu 1270
3503 GTATATGCGCATAGCGGTTTATGATTTTCCGCGACCTCAACAGCTT 3552
1270 hrProSerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuVal 1286
3553 TTC.....GCCGTACAGAGCAATGTGACCGCGCTGTTTG 3587
1287 PheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAl 1303
3588 CGAAGACCGCGCGCAACGCTTTGGACAAACGCGCATCCGCGACCAAC 3637
1303 aSerProHisAsnAsnValAlrPcLysAlaThrTYrAsn...ThrArgA 1319
3638 ACTACGCTGCGAAGATTTCGCGCTACCGCGCAACAGCGACCTGCC 3687
1319 snAsnValThrThrAspAlaGlyAlaGlyPheGluInThr..... 1332
3688 CAAATCGGATGCGAAGAAACCTC.....CG 3713
1333 LeuThrGlyMetThrValGlyIleAspSerProAsnAspIleProGlu 1349
3714 CAGCGGCGCGCTGCGCATCTGTTTTCGCAACCGCGCGAAGAAACNT 3763
1349 yIleAlaThrLeuGlyAlaPheMetGlyTYrSerHisSerHisIleGlyP 1366
3764 TCGACGACGCGCATCGGCAACTCGCGCGCTGCGCGCGCGCTTTC 3813
1366 heAspArgGly.....GlyHisGlySerVal... 1374
3814 GGGCAATACGCGATCGGC..... 3831
1375 GlySerTYrSerLeuGlyGlyTYrAlaSerTrpGluHisGluSerGlyPh 1391
3832 .....AGGTTGCGACATCGGCATC.... 3849
1391 eTYrLeuAspGlyValValLysLeuAsnArgPheGluSerAsnValAlaG 1408
3850 .....AGCAGCGCGCGGCTTTAGCAGCGCGCATCTTCACAGCGC 3891
1408 lYlswetSerSerGlyGlyAlaAlaAsnGlySerTYrHisSerAsnGly 1424
3892 ATCGGAGGCAAAATCGCGCGCGCTGCTGATTAACGCGCATTCAGGACG 3941
1425 LeuGlyGlyHis.....IleGluThrG 1432
3942 ATACGCGCGCGGCTTTCGGCGATTCGGCATCGACCGTACCGCGCA 3991
1433 yMetArgPheThrAspGlyAsnTrpAsnLeuThrProTYrAlaSerLeu 1449
3992 CGCGTATTTTCGTCMAAAGCGGATTCACGC.....TACGAA 4029
1449 hcGlyPheThrAlaAspAsnProGluTYrHisLysLeuSerAsnGlyMetL 1465
4030 AACGTCATATGCGCACCGCGGCTTTCGTTCAACGCGTACCGN...GC 4076
1466 SerLysSerValAspThrArgSerLe.....TYrArgGluLe 1478

```

```

4077 GGGCATTAAGCAGATTTATTCATTCAAACCGCGCGCAACATNTTCATCA 4126
1478 uGlyAlaThrLeuSerTYrAsnMetArgLeuGlyAsnGlyMetGluIleG 1495
4127 CMCCTTATTTNAGCGTCTTACCGATGCGGCTTCGGGCAAAATCCGA 4176
1495 lUpProThrLeuLysAlaAlaValArgLysGluPheValAspAsnArg 1511
4177 ACACGCGTCATATAC...GCNGTATTTGGCTCAGATTTTC...GGCAAAAC 4220
1512 ValLysValAsnAsnAspLysAsnPheValAsnAspLeuSerGlyArgAr 1528
4221 CGGAGTGGGGAATGGCGGTAAACCGCGAATCAAGGTTTACGCTGT 4270
1528 gGlyIleTYrGlnAlaGlyIleLysAlaSerPheSer...ThrLeuS 1544
4271 CCNCTCAGCGCTGCGCGCGCGCAAGGCGCAACTGTGAAGCGCAACAGC 4320
1544 eArgLysHisLeuGlyValGly.....TYrSerHisGly 1554
4321 GCGGCGATCAAA.....TTAGGCTACCGCTGG 4347
1555 AlaGlyValGluSerProTrpAsnAlaValAlaGlyValAsnTrp 1569
seq_name: p1r2:D85912

```

```

seq_documentation_block:
  hypothetical protein yp1A [imported] - Escherichia coli (strain 0157:H7, substrain E
  C;Species: Escherichia coli
  C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
  C;Accession: D85912
  R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
  iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimianta, E.; Potamousis, K.; Apod
  Nature 409, 529-533, 2001
  A;Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
  A;Reference number: AB5480; MUID:21074935; PMID:11206551
  A;Accession: D85912
  A;Status: preliminary
  A;Molecule type: DNA
  A;Residues: 1-1528 <STO>
  A;Cross-references: GB:AE005174; NID:g12517083; PIDN:AG57760.1; GSPDB:GN00145; UMG
  A;Experimental source: strain 0157:H7, substrain EDL933
  A;Genetics:
  A;Gene: yp1A

```

```

alignment_scores:
  Quality: 354.00      Length: 1667
  Ratio: 0.460        Gaps: 80
  Percent Similarity: 46.191  Percent Identity: 19.376

```

alignment_block:

US-09-303-518D-651 x D85912 ..

Align seg 1/1 to: D85912 from: 1 to: 1528

```

148 AACTACCAATACATTCGCGACTTGGCCGAAATTAAGCAAGTTTGCAGT 197
124 AsnGlnAspValTYrLysGlyGlyIleSerAsnGlyThrThrIleLysG 140
198 CGGGGCGAAGATTTAGGTNTACACAAAGGAGGAGTGGTC.... 243
140 yGlyAlaSerArgValGluGlySerAlaAsnGlyThrLeuIleAsp 157
244 ..GGCAATCAATGACAAAGCCCG...ATGATTGATTTTGTGGGTG 288
157 lYglYserGlnIleValLysValGlnGlyHisAlaAspGlyThrThrIle 173
289 TCGGTACGCGCGTGGCGGATTTGGT..... 315
174 AsnLysSerLysSerGlnAspValValGlnGlySerLeuAlaThrAsnTh 190
316 .....GGCGATCAATATATTT.....GTGA 334

```

```
190 rThtlleasnlglylarygintyvalgluglnserThrvallutht 207
      ||| |||||
335 GCGTGGCATATACGGCGCTATACACAGTT..... 366
      |||||
207 hThtllellysasnlglylgluglnarvalTyrgluserAralaleu 223
      |||||
367 .....GATTTGGTGGGAGAGAGNATCCCGATCGACCG 404
      |||
224 AspThrtlllelglylThtlserleuasnserlySerThral 240
      |||
405 TTTTCTTACCAATTTGTAAGAAATATATTAAGCCTTACATTCAC 454
      |||||
240 alyasnhtlrllelyrserglylThtlserleuasnhtlrl 256
      |||||
455 ACCCTTACACGGCGATTATATGCGGCTTGCATTAATTTGTCACA 504
      |||
257 .....Ser 257
505 GATGACAGACCTGTGCAATGACAGT.....GACATGAGGGG 542
      |||||
258 SerSeraspvalillegluvalTyrserglylvalleuaspvalsergl 274
      |||||
543 GAATACCTATTCGATTAAGAAATATCCCGAGCGTCCGATCGGCT 592
      |||||
274 yglYthralathtasnvalThtln..... 282
593 CAGGACACCCTATTGGCGTTATGATGATACAAACAGCGGATTATCC 642
      |||
282 ..... 282
643 TACTCGGCGCATGGTTAATGGCGCAATACACAT.....ATCA 683
      |||||
283 Htsaspglyala...lleuulsthtasnhtasnhtvalalaaspa 298
      |||||
684 GCGTTCGGGAATTAATGGCGTANTAGTTG..... 714
      |||||
298 rglYthrasnsergluglyalapheserillehtasnhtvalalaaspa 315
      |||||
715 .....ACGGCGATGGCGCCATGCCACAGCACTATGCGCCT 750
      |||||
315 snValleuenglunsnlglylHsleuaspilleasnhtalalygl... 330
      |||||
751 ATGCGGATTCAGTGGCGGACGAGCGGTTGCCAATGTTATTATTA 800
      |||||
331 .....SerAlaasnlysthtllellely 338
      |||||
801 TGACAAACAAACATTAATGGCTGCTCAAGGAGTTTACAAACGGCT 850
      |||||
338 sasplysglythrtmetServalleuthrasnalalysalaaspa... 353
      |||||
851 ACCCTATTCGCGGAGGAAACGGTTCCAGCTGATACGCAAGATTTGG 900
      |||||
354 .....Thrtarglleaspsnlglylvalmet..... 362
      |||||
901 TTCTAGATGACATTTTACAGAGCGATACACATACCGTCTNTTTGAACC 950
      |||||
363 .....Aspvalalaglyasnhtalhtasnhtllelle 373
      |||||
951 GCGCAGTACGACATTTTCTTACATTCACAAACAGCGTACGGGTA 1000
      |||||
374 .....Thrtlnasnhtleasnhtnlygllyllealt 387
      |||||
1001 CGGTACAGAAACACGAAAGTNTCCATCCAAACCTT.....AAA 1044
      |||||
387 hTglYthrasnhtleasnserglylThtlnasnhtlelysserglylys 403
      |||||
1045 GTACAGACAGTCCGACTGTTGACGATCTTGAATGAAGTGAATAAGA 1094
      |||||
404 AlaAspThrtlllelleserserglySerArglnvalvalglulysas 420
      |||||
1095 ACAGAGTT.....TACCGGCGAGGGGCTGTTAATCAAGTACC 1129
      |||||
```

```
420 pglYthralallelglYserasnhtleaserAlaglyly..... 432
1130 GTCCAGAGTTAAACACGGTGAAACCTTTCTTTATCATTCAGCGAAC 1179
      |||||
433 .....SerleullevalYthrtgly 439
      |||||
1180 GGCAAATCATCTTATCAACACATCAACCAAGGCGCGGCTTTGTA 1229
      |||||
440 Gly.....llealhtsglyvalasnhtlnhtlthglYserAlale 453
      |||||
1230 TTTTGAAGTGATTTTACGCTCCGCTCAAAACAAACGAAGCTGGCAG 1279
      |||
453 vAlalaasn.....Thg 458
1280 GCGCGGCGGTTCATACGT.....CAACAGATACCGTT 1314
      |||||
458 YalaglythraspillegluglyTyrasnhtlyserlehtshthrtlle 474
      |||||
1315 ACTTGAAGTAACGCGGTGCAACGACCGCTGTCCAAATGCGCAA 1364
      |||||
475 Thr.....glyglYlualasnhtYrvalvalleuGlun...Asnth 487
      |||||
1365 AGCAGCGCTCAGCTTCACGCCAAAGGAGAAACCAAGCTCGATCAGC 1414
      |||||
487 rglYgluthtlthrvAlvalalavsthtserAlatalsasnhtlthrtleA 504
      |||||
1415 TGGCGGAGGTACAGTATTTGATCAGCAGCAGCAGCAGATTAAGCMAA 1464
      |||||
504 spAlaglyglYlysleullelevalglhtlnhtlglulalys..... 516
      |||||
1465 AAACAAACCTTTAGTGAATGGCTTGNTCAGCGGAGGATACGTCGA 1514
      |||||
517 .....Thrtaspthrtlar 521
      |||||
1515 ACTGAATCGGATATACGTTCAACCCGACAAACTATTTGGGCTTC 1564
      |||||
521 gleuasn.....A 524
      |||||
1565 GCGCGGCGCTTTGATTTAAACGGCGATTCGTTCCACCGTATTT 1614
      |||||
524 snlglylvalleuenglualaspolyglYlualalyshts...Val 539
      |||||
1615 CAATATCCGATCAGGCGGATGATGATGATGATGATGATGATGATGAT 1664
      |||||
540 GluglnhtlserglYglYalaleulle.....Alaserthrt 552
      |||||
1665 ATCCACGTTACCATTTACAGGAATGAACT.....A 1696
      |||||
552 rserglYthrtleuhtlgluglythrasnserTyrglYaspalaphetYrt 569
      |||||
1697 TTACACACGAGTGTGAATATATCAATGACTTAATTCACGCAAGAA 1746
      |||||
569 leatrgasnserglualalysasnhtvalvalleuenglunhtlglYserleu 585
      |||||
1747 ATGCGCTACAAAGGTTGTTGGCAGAAAGATACGCAAGCAAGCAG 1796
      |||||
586 ThrtvalvalhtlthrtglYseratrgalalvalasphtlrllehtasnhtl 602
      |||||
1797 GCGGCTCAACCTT..... 1809
      |||||
602 ylysmetaspvalTyrglYlysaspvalglYthrtvalleuasnhtl 619
      |||||
1810 .....GTTTACAGCGCGCGGACAGAGCGGACCCGCTGCTT 1848
      |||||
619 lythrglnhtlrllelyraseratrgalalvalasphtlrllehtasnhtl 635
      |||||
1849 TCGCGGGAACAAATTTAAACGCAACATCAGCAAGCAAGCAAGCAAG 1890
      |||||
636 GlyglYlYsglnhtlthrtvalTyrglYleualathtlunhtlunhtlgluse 652
      |||||
1891 .GCGCAACGTTTTCAGCGGCGGACAGCAGCAGCAGCAGCAGCAGCAG 1939
      |||||
652 rglYgluglnhtlvalaspglYglYserthtlnhtlYlys.....Thrtlst 667
      |||||
```

```

1940 TAGAAGCGGCTGTGTCACAAATGAGAGTATCCACAAGA..... 1980
      :::::::::: ::::::::::
667 leaanglyglthrGlnthrValGlnsnrlyglYsAlaIleasnThr 683
1981 GAATCGGTGTGGACAAACGACTGCATCCGACGTTTAAAGGGGAAA 2030
      :::::::::: ::::::::::
684 AspilevalSerGlyLeuGlnGlnIleMetAlaasnGlyThrAlaGluGlu 700
2031 TTTCATATTTCAGGCGGACGCGGTGATTTCCCGC..... 2067
      ::::::::::
700 yserIleleasnGlySerGlnIleValasnGlyGlyLeuAlaG 717
2068 ..... 2085
      ::::::::::
717 luasnSerValLeuasnAspGlyThrLeuAspValArgGluLysGly 733
2086 GGCATTGNCATTTGACCAATCAAGCCCAAGCAGTTTGTGTGCGCAC 2135
      ::::::::::
734 SerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuValAlaThrTh 750
2136 GCATCAAGCGCATCAATCTGTACAGCTGCGACTGGACN..... 2175
      ::::::::::
750 rArgAlaThrArgValThrGlyThrArgAlaAspGlyValAlaPheSerI 767
2176 .....GCTGACAAAT..... 2187
      ::::::::::
767 leGluGlnGlyAlaAlaAsnAsnIleLeuLeuAlaasnGlyGlyValLeu 783
2188 TGTGTGCAANAANCAATTACCGAGCAATGAAGTATGCTTCA..... 2229
      ::::::::::
784 ThrValGlnSerAspThrSerSerAspLysThrGlnValasnThrGlyGlu 800
2230 .....TTGACTAAGACAGCAGTCNTNAGCGCAGTGTNAGCTNNCCNATN 2272
      ::::::::::
800 yArgGluIleValIleThrLysAlaThrAlaThrGlyThrLeuThrG 817
2273 ACGTNNNTTAAANCTCNCGGCGTGNCCACTNAN..... 2310
      ::::::::::
817 LysIleGluGlnIleValIleGluGlyValAlaasnGluThrThrIleAsnAsp 833
2311 .....GGCAATCTTAGTGAATGGAGGATACA...CGTATACAGT 2348
      ::::::::::
834 GlyGlyIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrThrI 850
2349 CAGCCACAAACGCCAC.....CAAAAGCGCAACTTACG...C 2383
      ::::::::::
850 eaasnGluGlyGlyThrLeuThrValasnAspasnGlyLysAlaThrAspI 867
2384 TCGTGGCAATGCCCAAGCAACATTAAATCAAGCCACATTAAAGCGCAAC 2433
      ::::::::::
867 leValIlelnsnSerGlyAlaAlaIleGluGlnThrSerThrAlaasnGly... 882
2434 NCATCGGNTTCGGGCAAT.....GCTCATTTAATCTAAGCAACA 2474
      ::::::::::
883 IleGluIleSerGlyThrHisGlnIleGlyThrPheSerIleSerGlyAs 899
2475 CGCCGACAAAC.....GCGAGTCTGACGCTTTCG 2506
      ::::::::::
899 lleuAlaThrAsnMetLeuLeuGluIleasnGlyLysnLeuLeuValIleuA 916
2507 ACAAGCCTAAGCAAAACGTAAAGCATTCGCGACTCAAGCGCAATGTCCTC 2556
      ::::::::::
916 leGlyThrGluAlaIleAspSerThrValGlyLysGlyGly..... 929
2557 CTAGCCGTAAGCAGATATTCATTTTGAANAACGCGCTTAAAGCGGACA 2606
      ::::::::::
930 .....AlaMetGlnasnGlnGlylnAspSerAlaThrLysVa 942
2607 ACTGACCGCGACAGANACAGATTACACTTAAAGAC..... 2646
      ::::::::::
942 lAsnSerGlyGlyGlnIleGlyThrLeuGlyArgSerLysAspGluPheGlnA 959

```

```

2647 .....ACGGAATGACCGCTCCCTACAGCGCAATTA... 2679
      ::::::::::
959 la**AlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleVal 975
2680 .....GCGAATTT 2687
      ::::::::::
976 ThrAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerI 992
2688 AAACCTT..... 2694
      ::::::::::
992 uSerIleuMetThrProArgAspAsnValThrProValLysLeuGluGlyA 1009
2695 .....GACAGCGCACCATTAACCTCAATTCGCCCTAT... 2727
      ::::::::::
1009 laIleArgIleThrAspSerAlaThrLeuThrIleGlyAsnGlyValAsp 1025
2728 .....CGCCAGATCTCCAGCGCGCAACAGCGCAGAGT... 2763
      ::::::::::
1026 ThrThrLeuAlaAspLeuThrAlaAlaSerArgGlySerValThrLeuAs 1042
2764 .....TCAGACAGCGCGCGCGCGCTTCGCGC...CGTT 2794
      ::::::::::
1042 nSerAsnAsnSerCysAlaGlyThrSerAsnGlyGlyThrArgValAsnS 1059
2795 CCTATT.....TCGGTTACACCG 2814
      ::::::::::
1059 eTLeuLeuLeuAsnAspGlyAsnValTyrLeuSerAlaGlnThrAlaAla 1075
2815 CCAACTCGGTGAATCCCGTTTCAACAGCTGACGAGTAACGCGCAATT 2864
      ::::::::::
1076 ProAlaThrThrAsnGlyIleGlyAsnThrLeuThrIleAsn...GluLe 1091
2865 GACAGCTAAGAGAACATTCGCTTATGTGGAAGCTTCGCGTACCGAA 2914
      ::::::::::
1091 uSerGlySerGlyAsnProPheTyrLeuIleThrAsnValAlaGlySerArg 1108
2915 GCGCAAAATTAAGCTGGCGGAAGTCCGAAGCACTTACACCTGCGC 2964
      ::::::::::
1108 LysAspGlnLeuValValAsnAsnAlaIleThrGlyAsnPheLysIlePhe 1124
2965 GTCAACATTAACGGGCAACAGCCGTAAGCCTGATCAATGAGCGTATG 3014
      ::::::::::
1125 ValGlnAspThrGlyValSerProGlnSerAspAlaMetThrLeuVa 1141
3015 GGAAGGAAAGCAACAACACCGCTGCGGAACCTTAATTATTCACCGTGC 3064
      ::::::::::
1141 l.....LysThrGlyGlyGlyAspAlaSerPheSerLeuG 1153
3065 AAAAGCAA.....CACGTGATCGCGCGCGGTGCGCTTACCACTCATC 3108
      ::::::::::
1153 LysAsnThrGlyGlyPheValAsnLeuGlyThrTyrGluTyr...ValLeu 1168
3109 CGCAAAAGCGGC.....GAGTCCGCTGATATCCGTCAAAGACA 3152
      ::::::::::
1169 LysSerAspGlyAsnSerAsnThrPAsnLeuThrAsnAspValLysPro.A 1185
3153 AGAGCTTCCGACAAACTCGCGCAGGCAAGAACCAAAAAACAGCGGAAA 3202
      ::::::::::
1185 snProAspProAsn..... 1189
3203 AAGACAACGCGCAAAAGCTTGACGCGCTGATTCGCGCGCGGCGATGCC 3252
      ::::::::::
1189 ..... 1189
3253 GCCGAAAAGACAGAAAGCGTTGCCGAACCGCGCGCGAGCGGGA 3302
      ::::::::::
1190 .....ProAsn..... 1191
3303 AAATGCGGCATTTATCGAGGCGGAGGAAAGAAAAAGGGGTGCAGCGCG 3352
      ::::::::::
1191 ..... 1191
3353 ATAAAGACAGCGCNTTGGCAAAACGCGGAGAACCGGACCGCGGNT 3402

```



```

90  IleasnlelyglythrlnthrValGlnasntrycilylsalalileasnth 106
1302 AGACGTACCGTTACTTGAAAGTAACGGCGTGGCAACGACCGCTGT 1351
106 rAsp..... 107
1352 CCAAAATCGCAAGGCGACGCTGCAGCTTCAGCCAAAGGGGAAACCA 1401
108  ....IleValserglyLeuGlnleuGlnleuMetalaasnlyThrAlaGlu 122
1402 GGCCTGATCGAGCGTGGCGACGCTACAGTATTGATTCAGCAGCGACA 1451
123 GlySerIleIle..... 126
1452 CGATAAAGCAAAACAAACGCTTTAGTGAATCGGCTTGNTCAGCGCA 1501
127  ....AsnnglyGlySerGlnValValasnGlnGlyLeu..... 138
1502 GGGGTACGGTGCACGTGAATGCCGATATTCAGTTCAACCCCGCAAACTC 1551
139  ....AlaGluAsnSerValLeuAsnAsp..... 146
1552 TATTTCGCTTTCGGCGCGACGCTTGATTTAAACGGCATTCGCTTTC 1601
147  ....GlyGlyThrLeuAspValArgGluLysGly...se 157
1602 GTTCACCGTATTCAAAATACGATGAGGGCGGATGATGNNATCAAA 1651
157 rAlaThrGlyIleGlnGlnSerSerIleGlnValAlaLeuVal..... 170
1652 ATGCCCAACACATCCACCGTTACCATTAACAGGAATGA..... 1692
171  ....AlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGly 185
1693  ....AGATTACACACGAGAGTGGTAGAATATCAATAGACTTAA 1733
186 ValAlaIlePheSerIleGlnGlnGlnValAlaAlaAsnAsnIle..... 198
1734 TTACAGCAAAAGAAATTGCTACACGCGTTGTTGGCGAGAAATACGA 1783
199  ....L 199
1784 CCAAAACGAGCGGGCGCTCAACCTTTTACCAGCCCGCGAGAAAGC 1833
199  euleuAlaasnGlnGlyValLeuThrValGlnSerAspThrSerSerAsp 215
1834 CGCACCCGCTGCTTTCGCGCGAGAACAAATTAAACGGCAACATCAGCA 1883
216 LysThrGlnValAsnMetGlyLysArgGluIleValLysThrLysAlaIth 232
1884 AACAAACGCAAACTGTTTTCAGCGGACAGACCGACCGACGCTTACA 1933
232 rAlaThrGly.....T 236
1934 ATCATTTGGAGAGCGGGTGGTCAAAATGGAAGATATCCACAGAGACA 1983
236 hrThrLeuThrGlyGlyGlnIleValGlnGlyValAlaAsnGlnThr 252
1984 ATCGTGTGGGACACGACTGATCNACCGCAGCTTAAAGCGGAAATTT 2033
253 ThrIle.....AsnAspGlyGlyIleGlnThrValSerAlaAsnGlyGly 267
2034 CCATTT.....CAGGGCGGCGAGCGGTGATTTCCCGCA 2068
267 uAlaIleLysThrLysIleAsnGlnGlyGlyThrLeuThrValAsnAsp 284
2069 AT.....GTGCAAAAGTGAAGGCGATTGNCATTTGAGC 2103
284 snGlyLysAlaThrAspIleValGlnAsnSerGlyAlaAlaLeuGlnThr 300
2104 AATCAGCGCCCAAGAGTTTGGTGTGCGACCGCATCAAGCCATACAT 2153
::: |||:::::: |||:::
301 SerThrAlaAsnGlyIleGluIleSerGlyThrHisGlnItyrGlyThrPh 317
2154 CTGTACACGCTTCGACGTGACNGCTGTGACAAATTTGTGCAANNAANCA 2203
317 e.....Seri 319
2204 TTACGAGCATTAAGTGAATGCTTCATTGACTTAACAGACATTNAGCGGC 2253
319 IeSerGlyAsnLeuValaThrAsnMetLeuLeuGlnAsnGlyGlyAsnLeu 335
2254 ANTGTNAGCTNNCCNATNACGNTNNNTNAAANCTCNCGCGCTNCGNC 2303
336 LeuValLeuAlaGlyThrGluAlaArgAspSerThrValGlyLysGlyGly 352
2304 ACTNANAGCAATCTTACGCAATGCGGATACAGT..... 2340
352 yAlaMetGlnAsnLeuGlnAspSerAlaThrLysValAsnSerGlyG 369
2341 ....TATACGTCCAGCAACGCGCACCCCAAAACGGCAACCTTAGCCTC 2385
369 LysIleThrThrLeuGlnItyrSerLysAspGluPheGlnAlaLeuAlaArg 385
2386 GTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAACGGCAACNC 2435
386 AlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValItyrAlaGlyTh 402
2436 ATCGGNTTGGGGCAATGCTTCATTTAATCTAAGACACAGCCGCAACAA 2485
402 rLeuAlaAspAlaSerValSer.....GlyAlaItr 412
2486 ACGGAGCTGACGCTTCCGCAACGCTAAGCAACGTAAGCATTC 2535
412 hrGlySerLeuSerLeuMet...ThrProArgAspAsnValItyrProVal 427
2536 GCACCTCAAGGCAATGCTCTCCAGCATTAAGCAGATTCATTTTGA 2585
428 LysLeuGlnGlyAlaValArgIleThrAspSerAlaThrLeuThrLeuG 444
2586 AAACAGCGCGCTTACCGGACACTCAGCGCAGCAAGANACAGCATTAC 2635
444 yAsn.....GlyValAspThrThrLeuAla 453
2636 ACTTA.....AAAGACAGCAATGAGAGCGTCCGTACGAGC 2673
453 spleuThrAlaAlaSerArgGlySerValItyr..... 463
2674 GAATTAGCAATTTAAACCTTGACACGCCACCATT...ACACTCAATTTC 2720
464  ....LeuAsnSerAsnAsnSerCysAlaGlyThrSerAsnCy 476
2721 CGCCTATCGCAGCATGCTGACAGCGCGCAAAACCGGACAGTGTACACA 2770
476 sGluTyrArgValAsnSerLeuLeuAsnAspGlyAspVal..... 490
2771 CGCGCGCGCGCGCTTCGCGCGCTTCCTTATATC.....GTACACG 2814
491  ....TyrLeuSerAlaGlnThrAlaAla 498
2815 CCAACTTCGCTAGCAATCCGTTTCAACAGCGCTACGCTAAACGGCAATT 2864
499 ProAlaThrThrAsnGlyIleTyrAsnThrLeuThrThrAsn...Glu 514
2865 GAACNGTCAGGAACATTCGCTTTATGTCGGAAGCTTCGCGCTACGCA 2914
514 userGlySerGlyAsnPheTyrLeuHisThrAsnValAlaGlySerArg 531
2915 GCGACAAATTTGAAGCTGGCGGAAAGTTCCGAAAGNACTTAACCTTGGG 2964
531 LysAspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPheLysIlePhe 547
2965 GTCAACATATCGGCAACGACCGGTAAGCTTCGATTCATTCAGCTTAC 3014
548 ValGlnAspThrGlyValSerProGlnSerAspAlaMetThrLeuVal 564

```

```

3015 GGAAGGAAACACAAACCGCTGCGAAACCTTAATTCACCTGC 3064
      |||
      |||
564 L.....LysThrIglyIglyAspAlaserPhehrleug 576
      |||
      |||
3065 AAAAGGAA.....CAGTCGATCGCGCGGCGCTTACCACCTATC 3108
      |||
      |||
576 LysThrIglyIglyPheValaspleugIyThrIglyIy...Valleu 591
      |||
      |||
3109 CGCAAGACGCG.....GAGTCCGCGCTGATATCGGTCAAGAAACA 3152
      |||
      |||
592 LysSerIasprglYasnsersasThrIasnsleuthrsasnsVallys 606
      |||
      |||
3153 AGAGCTTCCGACAAACTCGCAGAGGACGAAACCAAAACAGCGCGAAA 3202
      |||
      |||
606 ..... 606
3203 AAGACAGCGCGAAAGCCTTGACGCGGTGATGGCGCGCGGCGATGCC 3252
      |||
      |||
606 ..... 606
3253 GCGGAAAGACGAAAGCGTTGCGGACCGCGCGCGGCGGCGGGA 3302
      |||
      |||
607 ...ProAsnProAsnProIleProAsnPro..... 615
      |||
      |||
3303 AAATGTCGCAATTATGCAAGCGCGGAGAAAGAAAGAGGGTCCAGCGG 3352
      |||
      |||
615 ..... 615
3353 ATAAAGACAGCGCTTGGCGAAACGCGGAGCGGAAACCGCGCGGNT 3402
      |||
      |||
616 .....LysProAsnProIleProAsnPro 622
      |||
      |||
623 ProIysProAsnProAsnProIys.....ProAsnPro 633
      |||
      |||
3453 GCGAGCGCGGCGGCAACCTGCAACCGGACCGGCGGCGGCTGATNACC 3502
      |||
      |||
633 OthrProAsnProIleProIleProIleProIleProIleProIlePro 649
      |||
      |||
3503 GTTATGCGCAATAGCGGTTTGAATTTCCGCCACGCTCAACAGCGTT 3552
      |||
      |||
650 .....SerThrAlaValLeuValMetAlaIleThrLeuProleuVal 664
      |||
      |||
3553 TTC.....GCCGTACAGACGCAATTCGACCGGCTTGGC 3587
      |||
      |||
665 PheAspAlaIleLeuAsnSerIleArgIleuArgLeuAsnIleMetLysAl 681
      |||
      |||
3588 CGAAGACCGCGGCAACGCTTGGACAGCGGACATCCGGAACACCAAC 3637
      |||
      |||
681 AserProIleAsnAsnAsnValIleProIleAlaThrIleAsnIleArgAsnAla 698
      |||
      |||
3638 ACTACCGTTCCGAAATTCGCGCGCTACCGGCAACCAACGACCTGGCG 3687
      |||
      |||
698 snValThrIleAspAlaIleValIleArgIleuIleThrIleThrIleMet 714
      |||
      |||
3688 CAAATGGGATGCGAAGAAAC.....CTCGGAGGCGGCGCGCT 3725
      |||
      |||
715 ThrValGlyIleAsp**ArgAsnAspIleProGlyIleThrIle 731
      |||
      |||
3726 CGGCAATCGGTTTTCGACACCGGACGAAACAAACATTCGACGAGCGGA 3775
      |||
      |||
731 uclIalalpheMetGlyTyrSerIleSerIleIleGlyPheAspArgGly 747
      |||
      |||
3776 TCGGCAATCGGCGCGGCTTGGCGGCGGCGGTTTTCGCGCAATACGCG 3825
      |||
      |||
748 .....GlyIleIleGlySerVal...GlySerIleTyrSer 756
      |||
      |||
3826 ATGCGC..... 3831
      |||
      |||
757 LeuGlyIleTyrAlaSerIleThrIleAsnIleGlySerIlePheTyrLeuAsnProI 773

```

```

3832 .....AGGTTCGACATCGGCATC.....AGCA 3853
      |||
      |||
773 yValValIleValLeuAsnArgPheIleYserIasnsValAlaIleGlyLysMetSers 790
      |||
      |||
3854 CGGCGCGCGGATTTTACAGCGCGGCAATCTTTCACAGCGGATCGGAGCGAAA 3903
      |||
      |||
790 erGlyIleAlaIleAsnGlySerIleYhisSerIasnsIleuIleGlyIleHis 806
      |||
      |||
3904 ATCGCGCGCGCGCTGCTGATACGCAATTCAGGCGGCGGATTCACCGCGCG 3953
      |||
      |||
807 .....IleGlyIleIleMetArgPheHis 814
      |||
      |||
3954 TTTGCGCGGATTCGCGCATGCAACCGTACATTCGCGGACGCGCGCTATTTCG 4003
      |||
      |||
814 rasprglYasnsThrIleAsnIleThrProIleThrIleSerIleThrIleGlyPheHis 831
      |||
      |||
4004 TCCAAAAGCGGATTCACCGCTACGAAAC.....GTCAATATC 4041
      |||
      |||
831 laAspAsnProGlyIleYhisIleSerIasnsIleuIleYserIleYserVal 847
      |||
      |||
4042 GCCACCGCGCGCTGCTGCTGCAACCGTACCGN...GCGGCGATTCAGGC 4088
      |||
      |||
848 AspThrArgSerIle.....TyrArgIleuIleValAlaThrIle 860
      |||
      |||
4089 AGATTATTCATTCAACCGCGGCAACATTCATTCACACGCTTATTNA 4138
      |||
      |||
860 uSerIleYasnsMetArgIleuIleAsnGlyMetIleGlyIleIleProIleu 877
      |||
      |||
4139 GCGGTGCTTACCGGATCGCGCTTGGCGGCAACCTGCGGACGCGCTCAAT 4188
      |||
      |||
877 ysaIalaIalaValIleArgIleuIlePheValIleAspAsnAsnArgValIleValAsn 893
      |||
      |||
4189 ACCGCTGATTCGCTGCTGAGATTC.....GGCAAAACCGGAGTGGCA 4232
      |||
      |||
894 SerIasprglYasnsPheValIleAsnTyrLeuSerGlyIleArgIleIleYrgl 910
      |||
      |||
4233 ATGGGCGGTAACCGCGCAATTCAGGTTTCAGGCTTCCTTCGCGGCTG 4282
      |||
      |||
910 nalaGlyIleLysAlaSerPheSerIle...ThrLeuSerGlyIleHisIleu 926
      |||
      |||
4283 CCGCGCGGCAAGCGCGCACTGGAAGCGGACGACAGCGGCGGCAATCAA 4332
      |||
      |||
926 lYalGlyIle.....TyrSerIleSerAlaIleGlyValGlu 936
      |||
      |||
4333 .....TTAGCGTACCGCTGC 4347
      |||
      |||
937 SerProIlePheAsnAlaValAlaIleGlyValIleAsnTyr 947

```

seq_name: p1r2:E95206

seq_documentation_block:

cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain C:Species: Streptococcus pneumoniae

C:date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:accession: E95206

R:tel:tel: H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf on, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. A:Reference number: A95000; M01D:2157209; PMID:11463916

A:Accession: E95206

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4776 <KUD>

A:Cross-references: GR:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR: A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1772

alignment_scores:

287 Sergiyvallys.....AsnTYr..... 292

541 Alaserala...SeralaserThrseralaSeralaserIlese 556

A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: B90674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1327 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA83785.1; PID:913359819; GSPDB:G000154
A:Experimental source: Strain O157:H7, substrain RMD 0509952
A:Genetics:
A:Gene: EC50362

```
alignment_scores:
```

Quality:	285.50	Length:	1542
Ratio:	0.418	Gaps:	71
Percent Similarity:	44.293	Percent Identity:	18.872

```
alignment_block:
```

US-09-303-518D-651 x B90674 . .

Align seg 1/1 to: B90674 from: 1 to: 1327

364 GTTGGATTGTCGCGAAGGAAGNATCCGATGACGACCGTTTTCCTTA 413
:::|||||::: |||::: |||
59 IleasphneglyglyslnasnglnserIleasp.....Tyr 70
414 CCAATTGTGAAAGCAATTAATTAAG.....CTGACA 448
:::|||||::: |||
70 TargllethrValThrAspAsnAlaAsnLeuValIleasnaIaThrAsp 87
:::|||||::: |||
449 ATTCAACCCCTTACAAACGGCGATTANCATTATGCCGCGTTTCATAAATTT 498
:::|||||::: |||
87 hnsrArgPro..... 90
499 GTCACAGATGCAGAACTGTGAAAATGACGATGACATGAGGGGAATAC 548
90 90
549 CTAATCGATAAAGAAATAATCCGAGCGTGTCCGATCGGCTCAGAC 598
|||||::: |||
91ArgleuthrLeuAlaSerIle 97
599 ACCACTATTGGCGTTTATGATGATGACAAACACGCGGCGATTATCCTACTCC 648
||| |||::: |||
98GlyGlyLeuAspIleThr 103
649 GCGCGATGCTTA...ATTGGCGC...AATACACATATGACGGGTTG 689
|||||::: |||
104 GlyIglyIysValThrIleasnglyProLeuasnphLeuLeuIysGlyTh 120
690 GCGA.....AATAATGGCGTANTTAGTTGAGCGGCAT... 723
||| |||::: |||
120 rglYpHeleuasnValSerasnAlaGlySerGluLeuTyrAlaAspSpl 137
724GTGCGCATCCCAACGACATATGGCCCTATG 753
137 eutyrgIuSerasnSerGlyMetArgHisAspArgGlyTyrPheAsnVal 153
754 CCGATTCAGGTGCGCGACGAGCGGTTGCGCAATGTTTATTAT... 801
::: |||::: |||
154 SerasnIglyIysIleHisValIysGlyThrSerArgLeuThrArgLe 170
802GACAAACCAACCAATTAATGGCTGC 826
170 uGlnGlyAsnValSerGlyGlnGlySerGlnValAsnSerGluThrPhe 187
827 TCAACGGAGTTTACAACAC...GGCTACCCCTTATTCGGCAGGAA 870
::: ||| |||::: |||
187 hMetGlyValTyrGlySerTyrGlyIysnglnItyrLeuSerValAsn 203
871 AACGGTTTCCAGCTG...ATACGCAAAAGAT...TGTTTACGA 908
|||||::: |||
204 AsnclYglyIuValAsnAlaIArgIysGlnIleSerLeuclYtyrTyrAs 220

```

1  ::::::::::::::::::::
508 yThrValasphegIyala.....L 515

1622 CCGATGAGGGGCGATGATGNCNATATATATGCCACAAACATCCACC 1671
      |||||:::
515 eutThrAlaGlyAlaGlnMetThrGluGlyThrValAsnValSerIythr 531
      ::::
1672 GTTACCATTAACGGAATGAAATGATTACACACCGAGTGTGAATAT 1721
      ::::
532 LeuAspLeuArgGly...GluGlyValIleGlnValSerAspSerAspVa 547
      ::::
1722 CAATGACATTAATACAGCAAAATGCTTACACCGTGTGGTGGCG 1771
      |||:::
547 lValArg...SerValSerArgAspIle...AspSerAlaLeuSerL 561
      ::::
1772 AGAAAGATACAGCAAAACGAGCGGCGTCAACCTTTTACCAGCCC 1821
      ::::
561 eutThrGluValAspAspGlyAsnSerThrIleLysLeuValAspAlaGln 577
      ::::
1822 GCCGAGAGAGACCGCACCCNGCTGCTTCCGGCGGACAAATTTA.... 1866
      ::|||:::
578 GlYAlaGln.....ValLeuGlyAspAlaGlyAsnLeuGlnLe 590
      ::|||:::
1867 .....AACGCAACATC.....ACGC 1882
      |||||:::
590 uGlnAspLysAsnGlyGlnIleLeuSerSerSerAlaGlnArgAspIleG 607
      |||||:::
1883 AAACAACGCGCAACTGTTTTCAGCGC.....AGCCGACA 1920
      |||||:::
607 lGlnAlaGlnGlyGlnAlaValAlaGlyThrArgAspLysArgLeuThr 623
      ::::
1921 CGCAGCGCTACAAAT.....CATTTAGGAACGGGTGTGTCAAAAT 1961
      ::|||:::
624 SerGlyValAsnAsnAspGlyLeuIyThrIleGlyIyThrGlyLeuThrGlnLe 640
      ::|||:::
1962 GGAAGGTATCCCAAGAGAAATCGTGGGCAACAGCATGATCNAAC 2011
      ::|||:::
640 uAsp..... 641
      ::|||:::
2012 GCACGTTTAAACGGAAATTTCCATTTACAGGGGCGGCGGTGAT 2061
      ::|||:::
642 .....LeuHisAlaThrAspSerAspAlaLeuVal 651
      ::|||:::
2062 TCCCGCATGTGTCCAAAGTGGAGGCGATGNCATTTGAGC..... 2103
      ::|||:::
652 LeuSerSerAsnGlyLysSerGlnAsnAlaAlaAspLeuSerAlaLysI 668
      ::|||:::
2104 ....AATCAGCCCAAGCAGTTTGGTGTGCACCGCATCAAGCCATA 2149
      ::|||:::
668 eThrGlySerGlyAspLeuAlaPheSerSerGlnLysGlnIyThrValS 685
      ::|||:::
2150 CAATCTGTACAGT...TCGACGTGACNGGTGTGCAAAATTTGTGTGCA 2196
      ::|||:::
685 erLeuSerLysAspAsnAspLysIyThrGlyValThrAspLeuArgSer 701
      ::|||:::
2197 NAAANC.....ATTACGACGATTAAGTATGCTTCATTTGACTAAGAC 2240
      ::|||:::
702 GlyThrLeuLeuLeuAsnAsnAspAsnValLeuGlyAsn..... 714
      ::|||:::
2241 NGACNTNACGGCANTGTNAGNCTNNCATNACGNTNNTTNAANCNTCN 2290
      ::|||:::
715 .....ThrHisGluLeuArgLeuAlaAlaGluThrGluLeuAspMetA 729
      ::|||:::
2291 CNGCGCTCNCNCTNNAANGCATCTTATGCAATGGCATACAGCT 2340
      ::|||:::
729 snGlyHisSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeu 745
      ::|||:::
2341 TATACAGTACGCCACACGCCCAAAAGGCAACCTTAGCCTGTGGG 2390
      ::|||:::
746 .....LeuSerLeuAsnGln 750
      ::|||:::
2391 CATTGCCCAAGCAACATTTATTCAGCCACATTTAAAGGCAACNCATCGG 2440
      ::|||:::

```

```

750 yGlySerLeuThrValThrAsnGlyIyThrSerThrGlySerLeuThrG 767
      ::|||:::
2441 NTTCGGGCAATGCTTTCATTTATCTAAGCAACACCGCCCAAAACGGC 2490
      ::|||:::
767 LysGlyGluLeuAsnIle.....GlnGlyGly 776
      ::|||:::
2491 AGCTGACGCTTCCGACACGCTAAGCAACGTAAAGCCATTCGCACCT 2540
      ::|||:::
777 ThrLeuAspIleAlaGlyAsp.....AsnSerAsnLe 787
      ::|||:::
2541 CAACGGCAATGTCTCCCTAGCCGATAGCA.....GTATTCATT 2581
      ::|||:::
787 uThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerHisA 804
      ::|||:::
2582 TTGAAAACAGCGCCTTTACCGCACACTACGCGGACGAAGANACAGCA 2631
      ::|||:::
804 lAlnGly..... 806
      ::|||:::
2632 TTACACTTAAAGACAGCAATGACGCTGCCGTACAGCAGGAATTAG 2681
      ::|||:::
807 .....LeuG1 808
      ::|||:::
2682 CAATTTAAACCTTGAC...AAGCCACATTAACACTCAATTCGCTATC 2728
      ::|||:::
808 ySerAlaAsnValGluAsnAsnGlyThrLeuAlaLeuAsnAsnSerAlaG 825
      ::|||:::
2729 GCCACGATGCTGCGAGCGGCCAAACCGGAGNGTGTACAGACGCCGCC 2778
      ::|||:::
825 lLysAlaGlnAlaAlaSerValAsnIyThrAlaLeuGlyGlnLeuThr 841
      ::|||:::
2779 CGCGGTTCGCGCGCTTCCATTATTCGTTACCGCCCAACTTCGTTACA 2828
      ::|||:::
842 AsnAsnGly.....ThrLeuMetThrGlyMetSerGlyG1 853
      ::|||:::
2829 ATCCCGTTTACACAGCTGACGCTAAACGGCAATTGAAAC...NGTCAAG 2875
      ::|||:::
853 nGlnAlaGlyAsnValLeuValAlaLysGlyAsnIyThrHisGlyAsnAsnG 870
      ::|||:::
2876 GAACATTCGCTTATGTGCGAACTCTGCGCTACCGAAGC.....GAC 2919
      ::|||:::
870 lGlnLeuValMetAsnThrValLeuAsnGlyLysAspSerValThrAsp 886
      ::|||:::
2920 AAATTTGAAGCTGGCGGAAATTCGGAAGNACTTAACCTTGGCGGTCA 2969
      ::|||:::
887 LysLeuValAlaGluGlyAspThrSerGlyThrThrAlaValThrValAs 903
      ::|||:::
2970 CAAT.....ACCGCAACGAAACCGCTGACCTTGATCAATTGACG 3010
      ::|||:::
903 nAsnAlaGlyGlyThrGlyAlaLysThrLeuAsnGlyIleGluLeuIleH 920
      ::|||:::
3011 TAGTGAAGGAAAGACACAAACCGCTGCCGAAACCTTAATTTCACC 3060
      ::|||:::
920 lValAspGlyLysSerGluGlyGluPhe..... 929
      ::|||:::
3061 CTGCAAAACGAAACGCTGATGCGCGGCGTGGCGTACCACTCATCCG 3110
      ::|||:::
930 ValGlnAlaGlyArgIleValAlaGlyAlaIyThrAspIyThrLeuAlaAr 946
      ::|||:::
3111 CAAAGACGCGCAGTTCGCCCTGCATTAATCCGGTCAAGAAACAGAGCTTT 3160
      ::|||:::
946 gGlyGlnGly..... 949
      ::|||:::
3161 CCGACAAACTGCGCAAGGACGACCAACCAAAACAGGGGGAAGAACAC 3210
      ::|||:::
949 ..... 949
      ::|||:::
3211 GCGAAAGCCTTGAAGCGGCTGATGCGCGGCGGCGATCCGCGGAAA 3260
      ::|||:::
950 AlaAsnSerGlyAsnIyThrIyThrSerGlySerIyThrAspSerProGlu.. 965
      ::|||:::
3261 GACAGAAAGCGTTGCCGAACCGCGCGGCGAGGCGGGAATGTGCG 3310
      ::|||:::
965 ..... 965
      ::|||:::

```

3311 GCATTATGAGCGGAGAGAGAAAAACGGGTGACGGCGATTAAGAC 3360
 966LeuGlnProGluProAsp 971
 3361 AGGCGATTGGCGAAGACCGCGAAGGAAACCGCGGNTATACACCGC 3410
 972ProMetProAsnPr 976
 3411 CTTCGCCGCGCGCGCGCGCGCGGATTTCCGCGAACCACCGACCGC 3460
 976 ogLupProAsn.....ProAsnProGluProA 985
 3461 AACCCGAACTCAACCCCAACCGCAGCGACCTG..... 3495
 985 snProAsnProThrProThrProGlyProAspLeuAsnValAspAsnAsp 1001
 3496 ATAAAGCGCTTATGCCAATAGCGGTGAGTGAATTTCCGCCACGCTCAA 3545
 1002 LeuArgProGluValGlySerThrLeuAlaAsnLeuAlaAla...As 1017
 3546 CACGCTTTCCGCC.....GTACAGAGCAATTTG..... 3573
 1017 nThrMetPheThrThrArgLeuHisGlyArgLeuGlyAsnThrTyrTyrT 1034
 3574 ..GACGCGCTGTTCGCGAAGACCGCGCGACGCGTTTGACAAAGCNGC 3621
 1034 hAspMetValThrGlyGlnGlyGlnThrThrMetTrp..... 1047
 3622 ATCCGCGNACACCAACACTACCGTTCCGCAAGATTTCCGCGCTACCGGCA 3671
 1047 1047
 3672 ACAAAACGACCTGCGCGCAATCGTATGACGAAACCTCGCGACGGCGC 3721
 1047 1047
 3722 GGGTGGGATCTCTGTTTTCGCAACCGGACGGAACACTTCGACGAC 3771
 1048MetArgHisGlnGlyGlyHisAsnLysTrpArgAsp 1059
 3772 GGCATCGGC.....AAGTCGCAACGGCTTCCCGACGCG 3803
 1060 GlySerGlyGlnLeuLysThrGlnSerAsnArgTyrValLeuGlnLeuG 1076
 3804 C.....GCCGTTTTCGCGCAATACGCAATCGCAGGTTTCGACATCG 3844
 1076 yGlyAspValAlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValG 1093
 3845 GCATCAGCAGCGCGCGGGTTTACG...AGCGCAANTCTTCAGACGCG 3891
 1093 lValMetAlaGlyTyrGlyAsnSerAspSerLysThrLysSerArg 1109
 3892 ATCGGAGCAAAATCCGCGCGCGGTGCATTC.....GGCATTC 3935
 1110 ThrGlyTyrArgAlaLysAlaSerValAsnGlyLysThrGlyLeuTyr 1126
 3936 GGCACGATACCGCGCGGTTCGCGGATTCGCGCATCGAACCGTACATCG 3985
 1126 lAlaThrTrpTyrAla.....AspAspLysSerArgAsnG 1138
 3986 GCGCAGACCGCTATTTCTCCAAAAGCGGATTAACCGCTACGAAACGCTC 4035
 1138 lAla.....TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsn 1152
 4036 AATATCGCACCGCGCGCTTCGCTTCACGNTACCGN..... 4074
 1153 ThrValLysGlyAspAspLeuGlnSerGlnSerTyrLysSerLysGlyPh 1169
 4075GCGGGCATTAAGCAGATTAATTCATTCGAAACCGCGCAACACATNT 4120
 1169 eThrAlaSerLeuGlnAlaGlyTyrLysHisLysLeuAlaGlnLupheAsnG 1186

4121 CC.....ATCACNCTTATTNAGCTG 4143
 1186 lYSerGlnGlyThrArgAsnGlnTrpTyrValGlnProGlnAlaGlnVal 1202
 4144 TCCTATACCGATGCC..... 4158
 1203 ThrTrpMetGlyValLysAlaAspLysHisArgGlnSerAsnGlyThrLe 1219
 4159GCTTCGGCGCAAGTCCGAAACGACGCTCAATACCGCNG 4195
 1219 uValHisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysT 1236
 4196 TATTGGCTCAG.....GATTTCGCGCAAAACCGC..... 4224
 1236 hTrpLeuLysSerHisLysMetAspAspGlyLysSerArgLuphe 1252
 4225AGTCGGAATGGGCGCTTAACCGCGCAATCAAGGTTT 4262
 1253 GlnProPheValGluValAsnTrpLeuHisAsnSer.....LysAspPh 1267
 4263 CACGCTTCCTCCACGCTGCGCGCGCCCAAGNCCGCACTGGAAGCGC 4312
 1267 eSerThrSerMetAspGlyValSerVal.....ThrGlnAspGlyAla 1282
 4313 AACACAGCGCGGCGATCAATTAATGAGC 4338
 1282 rGAsnIleAlaGlnLupheLysThrGly 1290
 seq_name: p1r2:E85524
 seq_documentation_block:
 probable beta-barrel outles membrane protein 20402 [imported] - Escherichia coli (str
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: E85524
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Llim, A.; Dialianta, E.; Potamousis, K.; Apod
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
 A:Reference number: AB5480; MUID:21074935; PMID:11206551
 A:Accession: E85524
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1349 <SMD>
 A:Cross-references: GB:AE005174; NID:912513130; PIDN:AA654657.1; GSPDB:GN00145; UWCP
 A:Experimental source: strain 0157:H7, substrain EDL933
 C:Genetics:
 A:Gene: 20402
 alignment_scores:
 Quality: 285.50 Length: 1542
 Ratio: 0.418 Gaps: 71
 Percent Similarity: 44.293 Percent Identity: 18.872
 alignment_block:
 US-09-303-518D-651 x E85524 ..
 Align seg 1/1 to: E85524 from: 1 to: 1349
 364 GTGATTTTGGTGGCGAAGAGAAATCCCGATCAGACCGTTTCTTA 413
 81 lLeAspPheGlyLysGlnAsnGlnSerLLeAsp.....Tyr 92
 414 CCAAAATTGAAAAAATAATTATTAAG.....CCYGACA 448
 92 rArgIleThrValThrAspAsnAlaAsnLeuValIleAsnAlaThrAspT 109
 449 ATTACACCCCTTACACGCGGATTAACATATGCGCGTTTGCAATAATT 498
 109 hSerArgPro..... 112
 499 GTCACAGATGCAGAACCTGTGAAATGACGAGTACATGAGGGGGAATTC 548

112 112
549 CTATTCGATTAAGAAAAATATCCGAGCGTGTCCGATCGGCTAGAC 598
113 ArgLeuThrLeuAlaSerGly 119
599 ACCACTATTGGCGTTATGATGATGACAAACACGGCCGATTATCTACTCC 648
120 GlyGlyLeuAspLeuThr 125
649 GGGCAGTCGTAA...ATTGGCGC...AATACATATGCGGSGTGG 689
126 GlyGlyLysValThrIleAsnGlyProLeuAsnPheLeuLeuLysGlyThr 142
690 GGGA.....AATAATGGCGTAAATTAGTTGAGCGGCAT... 723
142 rGlyPheLeuAsnValSerAsnAlaGlySerGluLeuThrAlaAspAspL 159
724GTGGCGCATGCCACGACTATAGCGGCTATG 753
159 eutyrgLysAsnSerGlyMetArgHisAspArgGlyThrPheAsnVal 175
754 CGGATTGGAGGTGGCGGAGCGAGCGGTCGCCAATGTTTATTAT... 801
176 SerAsnGlyGlyLysIleHisValLysGlyThrSerArgLeuThrTyrl 192
802GACAAAACAACATTAATAGCGTCC 826
192 uGlnGlyAsnValSerGlyGluGlySerGlnValAsnSerGluThrPheP 209
827 TCACGGAGGATTTTACAAC.....GGCTACCGCTTATTCGCGAGGAA 870
209 hemetGlyValTyrgLysSerTyrgLysGlyAsnGlnTyrlLeuSerValAsn 225
871 AACGGTTTCCAGCTG...ATACGCAAGAT.....TGCTTCTACGA 908
226 AsnGlyGlyGluValAsnAlaArgLysGlnIleSerLeuGlyTyrlTyAs 242
909 TCACATTTCAGAGCGGATACACATACCGTCTNTTTGAA.....CCGC 952
242 pAlnVal.....SerAspThrThrLeuAlaValSerGluGlyGlyLysI 257
953 GCAGTAAGGACATTTTCTTACATCCAAACACAC..... 990
257 leSerAlaProThrIleSerLeuSerThrAsnSerGluLeuAlaLeuGly 273
991 ...GGTACCGGTACGTTACAGAAACCAACGAAAGGNTCCATCCAA 1037
274 AlaGlnGluGlySerAlaIleAlaLysAlaIleGlyIleIleAspAlaGlu 290
1038 GCTTAAGTACAGACAGTCGACTGTTGAGAA.....TCCTTGAATG 1081
290 sTleGluPheValThrAlaLysThrSerGluLysLysIleThrLeuAsnH 307
1082 AACTGATTAAGAACACAGTTTACGGCGCAGGGGCTTAATCAGTACCGT 1131
307 iSThrAspLysAspAlaThrIleSerAlaAspIleVal..... 319
1132 CCAAGGTTAAACAGCGTAAAACTTTCTTTATCGATTACGGCAAGG 1181
320SerGlySerGluGlyLeuGlyTyrlIleAsnAlaLeuAsnG 333
1182 CAACACTCATCTATCAAAACACATCAACACAGCGGCGGCTTGTGATT 1231
333 yThrThrTyrlLeuThrGlyAspAsnSerAlaPheSerGlyLysValLysI 350
1232 TTGAAGTGAATTTTACGCTCGCGCTGAACAAAC.....GAA 1269
350 leGluLinsnGlyAlaLeuGlyIleThrGlnAsnIleGlyThrAlaGlu 366
1270 ACGTGGCAAGGCGGCGCTTCAATATCACTGAGACAGTACCGTTACTTG 1319

367 IleAsnAsnArgGlyLysLeuHisLeuLysAlaAspSerMetThrPh 383
1320 GAAAGTAACGGCGGTGGCAACAGCACCGCTGTCCAAATCGCAAGGCA 1369
383 e...AlaAsnLysIleSerGlyAsnGlyThrIleSerIleAspSerGly 399
1370 CGCTGCAGCTTCACAGCCAAAGGGGAAAC.....CAAGCTCGATC 1410
399 hr.....ValGluLeuThrGlyAsnAsnTyrlAlaPheSerGlyTyrlIle 413
1411 AGCGTGGCGGAGGTACAGTCACTTTTGATCAGCAGGACAGCAGTAAGG 1460
414 AspValAlaSerGlyAlaValAlaIleSerGluAspLysAsnIleG 430
1461 CAAAAAACAGCGCTTAGTAATCGGCTTG..... 1491
430 yArgAlaGluLeuAspValAspGlyLysLeuGlnIleAsnAlaAsnLysA 447
1492NTCAGCGCAGGCGTACGCTGCACACTGAT 1521
447 sPTrValPheAspAsnAspLeuGluGlyArgGlyIleValGluIleAsn 463
1522 GCGGATTAAT...CAGTTCAACCCCGCAACACTTATTC..... 1557
464 MetGlyAsnHisGluPheSerPheAspGluPheAlaTyrlThrAspTrpH 480
1557 1557
480 eGlnGlySerLeuAlaPheGlnAsnThrThrPheAsnLeuGluLysAsnA 497
1557 1557
497 laGluPheLeuGlnLysGlyGlyIleThrAlaGlyGlnGlySerLeuVal 513
1558GGCTTCGCGGCGG 1571
514 ThrValGlyLysGlyAlaHisSerIleSerThrLeuGlyPheSerGlyL 530
1572 ACGTGGATTTAAAGCGGCTTCGCTTCGTTCCACCGTATTCAAAATA 1621
530 yThrValAspPheGlyAla.....L 537
1622 CCGATGAAGGCGGATGATGNCNATCATTAATGCCAACACATCCACC 1671
537 eutThrAlaGlyAlaGlnMetThrGluGlyThrValAsnValSerLysThr 553
1672 GTTACATTACAGGAATGAATATTAACAACAGAGTGAAGTAATAT 1721
554 LeuAspLeuArgGly...GluGlyValIleGlnValSerAspSerAspVa 569
1722 CAATGACTTAATTAATACAGCAAAAGAAATTCCTTCACAGGTGGTTGGCG 1771
569 lValArg...SerValSerArgAspIle...AspSerAlaLeuSerI 583
1772 AGAAGATACGACCAAAACAGACGGCGCTCAACCTTGTTCACGAGCC 1821
583 eutThrGluValAspAspGlyAsnSerThrIleLeuLeuValAlaPheAlaGln 599
1822 GCCGACAGAACCGCACCCGCTGCTTTCGCGGGAACAATTTA..... 1866
600 GlyAlaGlu.....ValLeuGlyAspAlaGlyAsnLeuGlnLe 612
1867ACGCGCAACATC.....ACGC 1882
612 uGlnAspLysAsnGlyGlnIleLeuSerSerAlaGlnArgAspIleG 629
1883 AAACAACGGCAACCTGTTTTCAGCGC.....AGACGACA 1920
629 lnsGlnAsnGlyGlnLysAlaIleValGlyThrAspTyrlArgLeuThr 645
1921 CCGCAGCGCTCAAT.....CATTTAGAACGCGGCTGTCAAAAT 1961
646 SerGlyValAsnAsnAspGlyLeuTyrlIleGlyTyrlGlyLeuThrGlnLe 662

```

1962 GGAAGTATCCACAGAGAGAAATCGTGTGGACAAAGACTGGATGCNAC 2011
      ::::
662 uasp..... 663
2012 GCACGTTTAAAGCGGAAAATTTCCATATTCAGCGCGGCGGCGGTGATT 2061
      ::::
664 .....LeuHisAlaThrAspSerAlaLeuVal 673
2062 TCCCGAATGTGCAAGTGAAGGCGATGTCATTGAGC..... 2103
      ::::
674 LeuSerSerAsnGlySerGluAsnAlaAlaAspLeuSerAlaLys11 690
2104 ....AATCAGCCCAAGCAAGTTTGTGTGCGACCGCATCAAGGCCATA 2149
      ::::
690 eThrGlySerGlyAspLeuAlaPheSerSerGlnGlyGlnThrValS 707
2150 CAATCTGTACAGT...TCGGACTGACAGCGTGTGCAAAATGTGTGAA 2196
      ::::
707 eThrLeuSerAsnLysAspAsnAspTyrThrGlyValThrAspLeuTrpSer 723
2197 NAAANC.....ATTACGACGATTAAGTATTGCTTCATTGACTTAAGAC 2240
      ::::
724 GlyThrLeuLeuLeuAsnAsnAspAsnValLeuGlyAsn..... 736
2241 NCACNTNAGCGGCANTGTNAGCNTNCCNATNAGCNTNNTTNAANCTCN 2290
      ::::
737 .....ThrHisGluLeuArgLeuAlaIleGluThrGluLeuAspMetA 751
2291 CNGGCGTGCNNCAGCTNANAGCAATCTTAAGTCAATGGCGATACAGCT 2340
      ::::
751 snGlyHisSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeu 767
2341 TATACAGTCAGCCACAGCCACCCAAAAGCGCACTTAAGCGCTGTGG 2390
      ::::
768 .....LeuSerLeuAsnGly 772
2391 CAATGCCCAAGCAACATTTAATCAAGCCACATTTAAAGCGCAACNCATCG 2440
      ::::
772 yGlySerLeuThrValThrAsnGlyGlyThrSerThrGlySerLeuThrG 789
2441 NTTCGGGCAATGCTTCATTTAATCAAGCAACAGCCCGCACAAAAGCGC 2490
      ::::
789 ySerGlyGluLeuAsnIle.....GlnGlyGly 798
2491 AGTCTGACGCTTTCGACACAGCCTAAGCCAAAGCTAAGCCATTCGGCACT 2540
      ::::
799 ThrLeuAspIleAlaGlyAsp.....AsnSerAsnIle 809
2541 CAAGCGCAATGTCTCCCTAAGCGGATTAAGCA.....GTATTCATTT 2581
      ::::
809 uThrAlaAsnValAlaIleAlaAsnSerAlaAsnValLeuValSerHisA 826
2582 TTGAAMAAGCGCGCTTACCGGACAACTCAGCGGAGCAAGANACAGCA 2631
      ::::
826 IagGlnGly..... 828
2632 TTACACTTAAAGACAGCGAAATGACGCTGCCGTACAGCAGCAATTAGG 2681
      ::::
829 .....LeuG1 830
2682 CAATTAAACCTTGAC...AAGGCACACATTACACTCAATCCGCCCTATC 2728
      ::::
830 ySerAlaAsnValGluAsnAsnGlyThrLeuAlaLeuAsnAsnSerAlaG 847
2729 GCCAGATGCTGACAGCGCGCAACCGGCAAGNGTGCAGACACCGCCGCG 2778
      ::::
847 IuLysArgAlaAlaAlaSerValAsnThrAlaLeuGlyGlyAsnLeuThr 863
2779 CGCGGTTGGCGCGCTCCCTATTTCGCTTACACCGCAACTTCGGTGAGA 2828
      ::::
864 AsnAsnGly.....ThrLeuMetThrGlyMetSerGlyI 875
2829 ATCCGCTTTCAACACGCTGACGCTAAACGGCAAAATTGAC...NETCAAG 2875
      ::::
875 nGlnAlaGlyAsnValLeuValValGlyGlyAsnTyrHisGlyAsnAsnG 892
2876 GAACATTCGCGCTTTATGTGCGAACTCTTGCGGCTACCGAAGC.....GAC 2919
      ::::
892 yGlnLeuValMetAsnThrValLeuAsnGlyAspAspSerValThrTrp 908
2920 AAATTGAAGCTGCGGAAAAGTTCCGAAAGNACTTACACTTGGCGGTCAA 2969
      ::::
909 LysLeuValValGluGlyAspThrSerGlyThrThrAlaValThrValAS 925
2970 CAAT.....ACCGCAACGAACCGGTAAAGCTGATCAATTGACGG 3010
      ::::
925 nasnAlaGlyGlyThrGlyAlaLysThrLeuAsnGlyIleGluLeuValLeu 942
3011 TAGTGAAGGAAAAGACAAACCGCGTCCGAAAACCTTAATTTCACG 3060
      ::::
942 lSValAspGlyLysSerGluGlyGluPhe..... 951
3061 CTGCAAAACGAACAAGTGCATGCGCGCGCGGTGCGCTTACCAATCATCG 3110
      ::::
952 ValGlnAlaGlyArgGlyValAlaGlyValAlaTyrAspTyrThrLeuAlaAr 968
3111 CAAGAGCGCGCAGTTCCGCTGCATTAATCCGCTCAAAAGACAGAGCTTT 3160
      ::::
968 gGlyGlnGly..... 971
3161 CCGCAAACTGCGCAGGAGAGAAAGCAAAAGCGGGAAGAGACAC 3210
      ::::
971 ..... 971
3211 GCGCAAAAGCTTGAACGCGCTGATTCGCGCGCGCGATGCCCGGAAAA 3260
      ::::
972 AlaAsnSerGlyAsnTyrPylLeuThrSerGlySerAspSerProGlu.. 967
3261 GACAGAAAGGTTGCGAACCAGCGCGCGCGCGAGCGGGAAGATGTCG 3310
      ::::
987 ..... 987
3311 GCATTATGAGGCGGAGAGAGAAAAACGGGTGCAGCGGATTAAGAC 3360
      ::::
988 .....LeuGlnProGluProAsp 993
3361 AGCCGNTTGGCGAAACAGCGCGGAAAGCGGAAACCGCGCGGNTACACCGC 3410
      ::::
994 .....PrometProAsnPr 998
3411 CTTCGCCCGCGCGCGCGCGCGCGCGGATTTGCGCAACCGCAGCGCC 3460
      ::::
998 ogLupProAsn.....ProAsnProGluProA 1007
3461 AACCGCAACTCAACCCCAACGCGACGCGACCTG..... 3495
      ::::
1007 snProAsnProThrProThrProGlyProAspLeuAsnValAspAsnAsp 1023
3496 ATNAGCGGTTATGCAATAGCGGTTGAGTATTTCCGCCACGCTCAA 3545
      ::::
1024 LeuArgProGluAlaGlySerTyrIleAlaAsnLeuAlaAla...As 1039
3546 CAGCGTTTTCGCC.....GTACAGAGCAATG..... 3573
      ::::
1039 nThrMetPheThrThrArgLeuHisLunArgLeuGlyAsnThrTyrTyrT 1056
3574 ..GACCGCGTGTTCGCGAAGACCGCGCAACGCGGNTTGGACAAAGCNC 3621
      ::::
1056 hTrsPheValThrGlyGlnGlyGlnLysGlnThrThrMetTrp..... 1069
3622 ATCCGGNACACCAACACTACGTTGCAAGATTTCGGGCGCTACCGGCA 3671
      ::::
1069 ..... 1069
3672 ACAACCGACCTGCGCCAAATGCTATGCAAGAAAAACCTGCGCAGCGGC 3721

```

```

1069 ..... 1069
3722 GCCTGCGCATCTGTTTCGACACGAGCGAAGAACATTCGACGAC 3771
      ::::: ::::: :::::
1070 ..... MetArgHisGluGlyGlyHisAsnLysTrpArgAsp 1081
3772 GGCATCGGC..... AACTCGGACGCGCTTGCCCGCAG 3803
      ||| ||| ||| ||| ||| |||
1082 GlySerGlyGlnLeuLysThrGlnSerAsnArgTyrValLeuGlnLeuG 1098
3804 C..... GCGGTTTCGGGCAATACGCGATCGCGGCTTGCAATCG 3844
      ||| ||| ||| ||| ||| |||
1098 YGlyAspValAlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValG 1115
3845 GCATCGACGACGCGCGCGGTTTACG...ACGCGCATNTCTGACAGCGC 3891
      ||| ||| ||| ||| ||| |||
1115 YValMetAlaGlyTyrGlyAsnSerAspSerLysThrIleSerSerArg 1131
3892 ATCGGAGCAAAATCCGCGCGCGCTGCTGATAC.....GGCATTC 3935
      ||| ||| ||| ||| ||| |||
1132 ThrGlyTyrArgAlaLysAlaSerValAsnGlyTyrSerThrGlyLeu 1148
3936 GGCACGATACCGCGCGGTTTCGCGGATCGCATCGAATCGATACATCG 3985
      ||| ||| ||| ||| ||| |||
1148 ValLeuThrTrpTyrAla.....AspAspGluSerArgAsnG 1160
3986 GCGCACGCGCTATTTCTCCAAAAGCGGATTCACGCTACGAAAAGCTC 4035
      ||| ||| ||| ||| ||| |||
1160 YAla.....TyrLeuAspSerTrpAlaGlnTyrSerTrpPheAspAsn 1174
4036 AATATCGCACCCCGGCTTGCCTTCAACCGNTACGNTACGNTACGNTAC 4074
      ||| ||| ||| ||| ||| |||
1175 ThrValLysGlyAspAspLeuGlnSerGlnSerLysSerLysGlyPh 1191
4075 ...GCGGCGCATTTAAGCAGATTTATTCATCAACCGCGCAACACATNT 4120
      ||| ||| ||| ||| ||| |||
1191 eThrAlaSerLeuGlnAlaGlyTyrLysHisLysLeuAlaGluPheAsnG 1208
4121 CC.....ATCACNCCTTATTATTCACCTG 4143
      ::::: :::::
1208 LysSerGlnGlyThrArgAsnGlnTyrTyrValGlnProGlnAlaGlnVal 1224
4144 TCCATATACGATGCC..... 4158
      ::::: :::::
1225 ThrTrpMetGlyValLysAlaAspLysHisArgGluSerAsnGlyThrLe 1241
4159 .....GCTTCGGGCAAGTCGGAACACGCGTCAATACCGCNG 4195
      ::::: ||| ||| ||| ||| ||| |||
1241 ValHisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysT 1258
4196 TATTGGCTAG.....GATTTCGGCAAAACCGC..... 4224
      ::::: ||| ||| ||| ||| ||| |||
1258 ThrTrpLeuLysSerHisLysLysMetAspGlyLysSerArgGluPhe 1274
4225 .....AGTCGCGAATGGGCGGCTTAACCGCAATCAAGAGTTT 4262
      ::::: ||| ||| ||| ||| ||| |||
1275 GlnProPheValGluValAsnTrpLeuHisAsnSer.....LysAspPh 1289
4263 CACGCTGTCNTCCACGCTGCGCGCGCAAGACCGCACTGGAAGCGC 4312
      ||| ||| ||| ||| ||| |||
1289 eSerThrSerMetAspGlyValSerVal.....ThrGlnAspGlyAla 1304
4313 AACACAGCCGCGGATCAATTAAGC 4338
      ::::: ||| ||| ||| ||| ||| |||
1304 rGAsnLleAlaGlnLysThrGly 1312

```

seq_name: p1r2:A43855

seq_documentation_block:
 high_molecular_weight_surface-exposed protein HKM1 - Haemophilus influenzae
 C:Species: Haemophilus influenzae
 C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

```

C:Accession: A43855
R:Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nonlysozyme
detella pertussis.
A:Reference number: A43855; MUID:92192797
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1536 <BAR>
A:Cross-references: GB:008876; GB:M84616; NID:9475770; PIDN:AAA20527.1; PID:9475771
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:89235, NCBIIP:89239)

alignment_scores:
  Quality: 277.00      Length: 1173
  Ratio: 0.474        Gaps: 58
  Percent Similarity: 49.872      Percent Identity: 20.375

alignment_block:
  US-09-303-518D-651 x A43855 ..

Align seg 1/1 to: A43855 from: 1 to: 1536

127 GGCACACCTTATTCGCGCATCACTACATGCGACTTGCCGA 176
    |||:::|||||:::|||||
349 GlyLutThrTyrLeuGly.....GlyAs 356
177 AATTAAGCAGATTTGCGACGCGCGGCAAGATATTGAGGNTACAACA 226
    ||| ||| ||| ||| ||| ||| |||
356 pGluArgGlyGlu.....GlyLysAsnGlyLysGlnLeuAlaLysL 370
227 AAAAAGGAGATTGTCGCGCAATCAATGACAAACCGCGATGATGAT 276
    || |||::: |||::: |||:::
370 YS.....ThrSerLeuGlnLysGlySerThrIleAsn 380
277 TTTTCTGTGTGTCGCGTAACGCGTGGCGGCACTTGCGGATCAATA 326
    ||| ||| ||| ||| ||| ||| |||
381 ValSerGlyLysGlnLysGlyArgAlaLysValTrpLysPheLea 397
327 TATTGTC.....AGCGTGCACATA 346
    ::::: ||| ||| ||| ||| ||| |||
397 AleuIleAspGlyAsnLleAsnAlaGlnGlySerGlyAspIleAlaLysT 414
347 ACGGCGCTAT.....AACAC 363
    ::||| ||| ||| ||| ||| ||| |||
414 hrgLysGlyPheValGlnThrSerGlyHisAspLeuPheIleLysAspAsn 430
364 .....GTGATTTTGGTGGCGAAGAGAAATCCGATCAGACCGTTT 407
    ||| ||| ||| ||| ||| ||| |||
431 AlaLleValAspAlaLysGlnTyrLeuLysProAspAsnValSerIle 447
408 TTCTTACCAATTTGTAAGAAATATTTATTAACCGTGAATTCACACC 457
    ::||| ||| ||| ||| ||| ||| |||
447 eAsnAlaGlnThrAlaGlyArgSerAsnThrSerGluAspGluTyrT 464
458 CTTCACACGCGGAT...TANCATATGCGCGTTTGCAATAATTTGTCACA 504
    ::||| ||| ||| ||| ||| ||| |||
464 hrgLysSerGlyAsnSerAlaSerThrProLysArgAsnLysGlnLysThr 480
505 GATCAGACACCTGTCGAATGACGAGTACATGAGGGGAATACCTATTC 554
    ::||| ||| ||| ||| ||| ||| |||
481 ThrLeuThrAsnThrThrLeuGlnSerIleLeuLysLysGlyThrPheVa 497
555 CGATTAAGAAATAATATCCGACGCTGTCGATCGCTCAGACACACAC 603
    ::||| ||| ||| ||| ||| ||| |||
497 LAsnIleThrAla...AsnGlnArgGlyTyrValAsnSerSerIleAsnL 513
604 .....TATTGGCGTTAT..... 615
513 euserAsnGlySerLeuThrLeuTrpSerGlnGlyArgSerGlyGlyGly 529

```

616GATGATGACAAACGGCGA 635
530 ValGluIleAsnAspIleThrThrGlyAspAspThrArgGlyAlaAs 546
636 TTTATCC...TACTCCGGCGCATGTTA.....A 661
546 nleuThrIleTyrSerGlyGlyTyrValAspValHisTyrAsnIleSerL 563
662 TTGGC.....GGCAATACACATATGACGGT..... 687
563 euGIAlaIleGlnIleAsnIleAsnIleThrAlaAlaGlnAspIleAlaPhe 579
688 ...TGGGGAATATATGGCGTANTTATGATGACGGCGCATGTGCGGCATGC 734
580 GluTyrSerAsnGlnValIleThrGlyGlnGlyThrIleThrSerG 596
735 CAACACATATGGCCCTATGCCATATGACAGTGGCGGCGACAGCGGTT 784
596 yAsnGlnTyrGlyPheArgPheAsnAsnValSerLeuAsnGlyThrGlyS 613
785 CGCCATGCTTTATTTATGACAAACAAACATTAATGGCTGCTCAACGGA 834
613 erGlyLeuGlnPheThrThrIleTyrArgThrAsnTyrAlaIleThrAsn 629
835 GTTTTACAAACCGCTACCTTATTCGGCAGGAAACGGTTTCCACT 884
630 LysPheGlnGlyThrLeuAsnIleSerGlyValAsnIleSerMetVal 646
885 GATACGCAAGATGTGCTCTACGATGACATTTACAGCGCATACACT. 933
646 lleuProLysGlnSerGlyTyrAspLysPheLysGlyArgThrTyrT 663
934ACCGTCNTTTTGACCGCGCATACGACGATTTTCCTTTCA 978
663 rPAsnLeuThrSerLeuAsnValSerGlnSerGlyLysPheAsnLeuThr 679
979 TCCAAACAACAACGGTACG.....GTCACGGTACAGAA..... 1011
680 IleAspSerArgGlySerAspSerAlaGlyThrLeuThrGlnProTyrAs 696
1012 .ACCAACGAAAGGTNTCCAACTCCAAAGCTTAAAGTACGACAGCTCCGAC 1060
696 nleuAsnGlyIleSerPheAsnLysAspThrThrPheAsnValGluArg. 712
1061 TGTTCGCAAAATCTTTGAATGAACGATGAAGAAGCAGTTACGGCGCA 1110
713AsnAlaArgValAsnPheAspIleLysAlaProIle..... 724
1111 GGGGTGTATACGTACGTCGCAAGGTTAAACAACGCTGAACCTTTC 1160
725 ...GlyIleAsnLysTyr.....Se 730
1161 TTTTATGATTAACGCAACGCAACATCTATTCAAACAAATCAAC 1210
730 rSerLeuAsnTyrAlaSer.....PheAsnGlyAsnIleSerV 743
1211 AAGCGCGGCGGCTTTGATTTTAAAGGTATTTTACGCTCCGCTGA 1260
743 aLserGlyGly.....SerValAspPheThrLeuLeuAlaSer 756
1261 AACACGAAACGTGCAAGCGCGCGGCT..... 1290
757 SererAsnVal...GlnThrProGlyValIleAsnSerLysTyrPh 772
1291 .CATATCAGTAAGACAGTACGTTACTTGAAAGTAAGACGC..... 1332
772 eAsnValSerThrGlySerSerLeuArgPheLysThrSerGlySerThrL 789
1333GTGGCAACGAC...CGCTGTCCAAATGGCGAAA 1365
789 ysthrGlyPheSerIleGlnLysAspLeuThrLeuAsnAlaThrGlyGly 805
1366 GGCACGCTGACGTTCAAGCCAAAGGGGAAAC..... 1398
806 AsnIleThrLeuLeuGlnValGlyGlyThrAspGlyMetIleGlyGly 822
1399CAAGCTCGATACAGCTGGCGGACGCGTACGATCATTTGG 1438
822 yIleValAlaLysLysAsnIleThrPheGlnGlyGlyAsnIleThrPhe. 838
1439 ATCAGCAGGCGACGATTAAGGCAAAACAGCTTTAGTGAATCGGC 1488
839GlySerArgLysAlaValThrGluIle... 847
1489 TTGNTCAGCGGCGGCTACGGTGCACATGAAATGCCATTAATCATTTCA 1538
848GluGlyAsnValThrIleAsnAsnAsnLysn..... 858
1539 CCCCAGAACCTATTTTCGCTTCCGCGGCGACGCTTGTGATTAACG 1588
859ValThrLeuIleG 863
1589 GGCATTCGCTTTCCTCCACCGTATTCAAATACC...GATGAAGGGCG 1635
863 LysrAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspVal 879
1636 ATGATTCNCATCATATGCGCACACA...ACATCCACGCTTACCATTC 1682
880 IleIleAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAl 896
1683 AGCGATGAAGATATTACACACCGAGTGGTGAAGATATC..... 1722
896 aGlyAsn.....LeuThrValGlnSerAsnAlaAsnPheLysAlaIle 911
1723ATAACACTTAATTC 1737
911 hrAsnPheThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSer 927
1738 AGCAAAAGAAATTCCTACACAGGT...TGGTTGGCGAGAAAGATACGAC 1784
928 AsnIleSerIleAlaLysGlyGlyAlaArgPheLysAspIleAspAsn 944
1785 CAAACGACAGCGCGCTACCTGTTTACAGCCCGCGCGCAAGAAC 1834
944 rLys.....AsnLeuSerIleThrThrAsnSerSerThrTyrA 958
1835 GCACCCNGCTGTTCCGGCGGACAAATTAAGGCAACATCCAGCA 1884
958 rghrIleIleSerGlyAsnIleThrAsnLysAsnLysPheAsnIle 974
1885 ACAACGCAAACTGTTTTCACGGCGCAGACGACACCGCAGCCCTACA 1934
975 ThrAsn.....GluGlySerAspThrGlu.....Me 983
1935 TCATTTAGAGCGGGGTGTCGCAAAATGAAGGTATCCCAAGAGGAAA 1984
983 tGlnIleGlyGlyAspValSerGlnLysGly..... 994
1985 TCCTGTGGGACACGACTGATCNCACGACGTTTAAAGCGGAAATTC 2034
995AsnLeuThrIleSererAspLysIle 1003
2035 CATATTCAGGCGCGGACGCGGATTTCCCGCAATGTTGCCAAATGGA 2084
1004 AsnIleThr...LysGlnIleThrIleLysAlaGlyValAspGlyLys 1019
2085 AGCGATTCGATTTGACGATACGACCCCAAGCAGTTTGGTGTGCGCAC 2134
1019 nSerAspSerAspAlaThrAsnAsnAlaAsnLeu..... 1030
2135 CGCATCAAGCCATACATCTGTACAGTTCGACTGACACNGTGTGACA 2184
1031ThrIleLysThrLys..... 1035
2185 AATGTGTGAAANAAACATTCACGACGATAAGATGCTTCAATTGAC 2234

1036 GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049
2235 TAAAGCAGNCAGCTNAGCGGCAGTGTAGNCCTNNCCATACGCTNNTTAA 2284
1049 nlysalagluilethrAlaLys 1056
2285 ANCTCNCNGGCGNTGCGNCCACTNMANNGCAATCTT... AGTCGAATATGC 2331
1057 AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071
2332 GATACACGTTATACACTGACGCCACACGCGCACCCAAAAGCGCAACTTAG 2381
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088
2382 CCGTGGGGCAGTATGCCACGAACAATTAAACCCCACTTAAACGGCA 2431
1088 TAlaAspGly..... HisLysValThrLeuHisSerL 1099
2432 ACNCATCGGNTTCGGCGCAATGCTTCATTTATCTAAGACACAGCGCGCA 2481
1099 ysalGluThrSerGlySerAsnAsn... AsnThrGluAspSerSerasp 1144
2482 CAATAAGCGCAGTCTGACCTTTCGCCACAACGCTAAGCAACGTAAGCA 2531
1115 AsnAsnAlaGlyLeuThrIle..... AspAlaLys... AsnValThr.. 1127
2532 TTCGGCAGCATACGCAAGATGTCCTCCATGACCATTAAGCAGATTCATCT 2581
1128 ValAsnAsnAsnIleThr... SerHisLysAlaVal 1138
2582 TTGAAAACAGCGCGCTTTACCGGACACATCAGCGCGACAGAAAGAACCA 2631
1139 .. SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyLysThr 1154
2632 TTACACTTAAAA..... GACAGCGAATGAGCGCTCGCGTCAGAGGACGA 2675
1155 IleAsnAlaThrThrGlyLysAsnValGluIleThrAlaGlnThrGlySerI 1171
2676 ATTAGCAATTTTAAACCTTGACAGCGCCACCTACACTCAATTCGCGCT 2725
1171 eleuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaThr.. 1187
2726 ATGCCACAGATGCTGAGCGCGCAACCGGAGNGTGTACAGACGCGCG 2775
1188 GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201
2776 CGCGCGCGTTCGGCGCGCTTCCTATTATCCGTATACCGCCCAACTTCGCT 2825
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrI 1218
2826 AGAATCCCGTTTAAACAGCTGACGCTTAACCGCAATTTGAACNCTCAG 2875
1218 elyGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGly 1235
2876 GAACATTC..... CGCTTATGCGGAACCTTCCT 2904
1235 LyrThrIleSerGlyThrValGluValLysAlaThrGlnLysLeuThr 1251
2905 GCGTACCGAAGCAGCAAAATTTGAAGCTGCGGGAAGTTCGGAAGACCTTA 2954
1252 ThrGlnSerAsnSerLysIleLys... AlaThrThrGlyGluAlaAsnVa 1267
2955 CACCTTGGCGGCTC..... AACATTCACGGCA 2980
1267 LThrSerAlaThrGlyThrIleGlyLysThrIleSerGlyAsnThrVal 1284
2981 ACGAACCGCTAAGCGTCGATCACTTACGCTA..... 3012
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyLysnGlyAlaGluIle 1300
3013 GTGGAAGGAAAGCAACAACGCGCTTCGCGAAACCTTAATTT 3056
1301 AsnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThr 1317

```

3057 CACCCTGCAAAAGCAAGCATGTCATCCGCGCGGTGACGTTACCACTCA 3106
      |||  :::::||||:::  :: ||:::||||
1317 rThrgtUaIaSerSeHiiStIeHrSer .AlaYsgIyGInVaIaSnIeU 1333
      |||  :::::||||:::  :: ||:::||||
3107 TCCGCAAAAGACGGCGAGTTCGCGCTGCATATCCGTCGTAAGAACAAGAG 3156
      |||||:::  ::|||  |||  |||:::||||  |||
1334 SerIaGInaSpIySerVaIaInGlySerIleasnIaIaIaSnVaIth 1350
      |||  :::::||||:::  :: ||:::||||
3157 CTTTCGCAACAACCTGGCGCAAGGCAAGAACCCAAAACAGCGCGCAAAAAGA 3206
      :::::  |||:::||||:::  :: ||:::||||
1350 rIeuasnrItrIthGlyrHrIeuthrThrVaIlySgIySerIasnIleasnA 1367
      |||  :::::||||:::  :: ||:::||||
3207 CAACGCGCAAAAGCCTTG 3223
      ::|||  :::::||||
1367 IaThrSerGlyThrIeu 1372
      |||  :::::||||
seq_name: pIrr2.G64993

```

```

seq_documentation_block:
  yfai_protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: G64993
R:Blatther, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
  A.: Rose, D.J.; Mau, B.; Shao, Y.
  Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
  Reference number: A64720; MUID:97426617
A:Accession: G64993
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1250 <R1A>
A:Cross-references: GB:AE000313; GB:U00096; NID:g2367132; PIDN:NA052293.1; PID:g17885
A:Experimental source: strain K-12, substrain MO1655
C:Genetics:
  A:Gene: yfai

```

```

Alignment_scores:
      Quality: 263.50      Length: 1438
      Ratio: 0.422      Gaps: 63
      Percent Similarity: 43.394      Percent Identity: 19.402

alignment_block:
US-09-303-518D-651 x G649933  ..

Align seg 1/1  to: G64993  from: 1  to: 1250

697 AATGGCGTANTAGTTTATGAGCGGC.....GATGTGGCCCA 731
|||||.....:|||||:
25 ASngIyVAIAIAIAValThraspsercysGInlyTyraspVAlYsAl 41
732 TGCACAGCGATATGGCCCTATCGGATTCAGTGCAGTGGCGGAGCGAC.... 777
|||||.....:|||||
41 asercysGInAlaseraTgInserleuSerGlyIleThrGInAsprIps 58
|||||.....:|||||
778 .....AGCGTTCCGCAATGTTTATTTATGCAAAACAACAACATATAA 819
|||||.....:|||||
58 exIleAlAspGlyInTrpLeuValPheSerAspMelThraspAla 74
|||||.....:|||||
820 TGGCTGCTCAACGGAGGTTTACCAAAACGGCTACCCCTATTTC..... 861
|||||.....:|||||
75 SerGlyGlyAlaValAlpheLeuGInGInGlyAlaGInUpheSerleuLeupr 91
|||||.....:|||||
862 .GGCAGGAAACGGTTTCCAGCTGATACGCAAAAGATTGGTTTACGATG 910
|||||.....:|||||
91 oGInAsnGInUrTrGlyMeTrhLeupheAlaAsnAspTrValThrGlyG 108
|||||.....:|||||
911 ACATTACAGAGC.....GAT 927
|||||.....:|||||
108 InTrYAsnAsnGlyAlaValAllePheAlaIAsnGInAsnSerThrLeuAsn 124
|||||.....:|||||

```


[illegible]


```
1029 1ag1yLeuGlnAsnSerTrpMetAsnPheIleValAsn..... 1042
1036 .....AAGCTTAAAGTACAGACAGTCCGACTGTTGACGAACTTTGA 1078
1037 yGlyPheAsnLeuValAsnAsnIle..... 1068
1079 ATGAAGTATGATTAAGAACGTTTACGGCGGCGGCTTAAATCAGTAC 1128
1069 ..ThTrpAspLysSerValSerGlyGlyGlyAsnPheGlyValAsp 1084
1129 CTTCCCAAGTTAAACAAGCT.....GAAACCTTCTCTTATATGA 1169
1085 AsnAlaAsnAlaAsnGlyAsnAlaValIleLysAsnValAsnPheSerAs 1101
1170 TTACGGCAACGGCAAACTCATCTTA..... 1194
1101 p.....AsnGlyThrLeuIleTyrLysGlyGlyGlnSerAlaGlyA 1116
1195 .....TCAACACATCAACCA 1212
1116 snSerLeuThrLeuGlnAsnAsnThrPheAsnSerTyrAsnIleAsnAla 1132
1213 GCGCGCGGCGGCTTGTATT.....GAAGGTGATT 1244
1133 LysAlaGlnAsnLeuIlePheAsnAsnAsnSerPheAsnSerGlySery 1149
1245 TACGCTCTGCTGAAACCAACGAAACGTGGCAGCGCGGCTTACATA 1294
1149 rSerPheAsnAspThrLysAsnValThrPheLysGlyThrAsnThrLeu 1166
1295 TCAGTCAACAC.....AGTACGTTACTTGGAAAGTA 1326
1166 leAsnSerAspProPheSerArgLeuLysGlySerValSerIleAspAsn 1182
1327 AACGGCTGGCAACGACCGCTGTCGCAAAATCGGCAACGACACCTGCA 1376
1183 AsnSerIlePheAsnIleGlyArgAspLeuThrAspLysThrThrTyrTh 1199
1377 CGTT.....CAAGCCAAAGGGGAAA 1396
1199 leuLeuSerGlyAspAsnIleLysTyrAsnAsnGlnAlaLeuAlaAspA 1216
1397 AC.....CAA 1401
1216 snValPheSerLysAsnLeuTrpAspLeuIleHisTyrAspGlyGln 1232
1402 GCGTCATGACGCGTGGCAGCGTACAGTCAATTTGGATCAGCAGCAGA 1451
1233 GlyThrLeuLeuArgThrAspAsnAsnThrTyrPheValGlnPheThrCl 1249
1452 CGATTAAGGCACAAAACAAGCTTTAGTGA..... 1482
1249 nSerAsnGlyGlnLysPheValPheGlnGlnThrPheAsnProGlySery 1266
1483 .....ATCGGCTGNTCAGCGGCGACGGGTACGGTCAACAG 1518
1266 leThrTyrLysTyrPheThrIleHisSerSerProPheHisThrGlnAla 1282
1519 AATGCGATTAATCAGTTCAACCCGCAAACTTATTTGGCTTTCGCGG 1568
1283 AspSerLysAspLleTrpAsnGlnValArgLysGlnPheAspPheIleTr 1299
1569 CGGACGCTTG..... 1578
1299 ogLysThrProValLysValGlyValLysTyrIleAlaProTyrLysA 1316
1579 .....GATTAAACGGGCAT.....TCGCTTTCGTTCCAC 1608
1316 snGlnAspLeuIleGlySerSerAlaPheAlaTrpSerLeuAsnPhe... 1331
1609 CGTATTCAAAATACCGATGAAGGGCGGATGATTGNCNATCATGATGCGAC 1658
1332 .....GlyAlaThr 1334
1659 AACAACTCCACCGTTACATTCACAGGGAATGAAGTATTACACAAACGA 1708
1334 rValValGlyThrLeuLeuLeuGlySerAlaGlnGlnLysAlaAsnAsnA 1351
1709 GTGTAAAGATATTCATAGACTTAATTACAGCAAAAGAAATGGCTTAAAC 1758
1351 snGlyGlySerIle..... 1355
1759 GGTGTGTTGGCGAGAAATACACCAACCAACGAGCGGCGCTCAACCT 1808
1356 ..ThrPheGlyLysAsnAsnLeuLeuTyrLeuHisGlyLysPheAsnAl 1371
1809 T.....GTTTACACCGCGCGCGAG 1828
1371 aThrAsnIlePheLeuThrAsnAsnPheAsnValGlyAsnProAsnAla. 1387
1829 AAGACCGCACCCNCGCTCTTCGGCGGCAACAAATTTAAACGGCAACATC 1878
1388 .....GlyGlyGlyAlaThrIleAsnPheAsnAla 1397
1879 ACGCCAACA..... 1887
1398 AspLysThrLeuSerAlaAspGlyLeuAsnTyrThrAsnPheGlnThrVa 1414
1888 .....A 1888
1414 lAlaMetGlyLeuGlnThrSerAlaSerGlnHisSerTrpAlaAsnPheA 1431
1889 ACGCCAAACTG..... 1899
1431 snSerLysLeuSerMetGlnIleLysAsnSerAsnPheArgAspPheThr 1447
1900 .....TTTTACGGCGCAGCAGCACACCGCGACCGCTTACAA 1934
1448 TrpGlyGlyPheArgPheAsnSerGlyArgIleThr.....PheGlnAs 1462
1935 TCATTTGAAGACCGGTGCTCAAAATGAAAGTATCCCAACAAAGA.... 1980
1462 nThrThrPheSerGlyTrpThrAsnIleAsnGlyAlaThrGlySerLys 1479
1981 .....GAATCGTGGGAGCAACGATCATCNCGCC..... 2013
1479 erSerTyrValAsnMetValAlaAsnThrAspLeuIlePheThrAspSer 1495
2014 .....ACGTTAAACGGGAAATTTCCATAT 2039
1496 lleLeuGlyGlyIleArgTyrAspLeuLysAlaAsnIleIlePhe 1512
2040 TCAGGCGGCGGAGCGGTG.....ATTCCCGCAATGTTGCCAAACTGG 2083
1512 eaAsnAsnThrClMetValAlaAspValSerLysAsnValAsnGlnSers 1529
2084 AAGCGATTCGATTTAGC...AATCAGCCCAAGCAGTTTGTGTGCT 2130
1529 erLeuAsnGlyAsnValThrPheAsnHisSerArg.....LeuSerVal 1543
2131 GCACCGCATCAACCATACATCTGTACAGTTCGAGCTGAGACNGCTCT 2180
1544 LysProAsnAlaIleAlaIleAsnIleGlyLysAspIleThrGlnThrThrLe 1560
2181 GACAAATTTGTCGAANAANACATTACCGACAGATTAAGTATGCTTCAT 2230
1560 uGlnAsnAlaSerSerLeuSerPheTyrAsnAspSerValAlaAsnPheA 1577
2231 TGACTAAGACGACNTNAGCGGCANTGTNAGNCNTNACNTNACGNTNNT 2280
1577 snGlyThrThrAlaPheAsnGlyValSerTyrLeu..... 1588
```

```

2281 TNAANCTCNNGGCGTGCNACTMAAGCAATCTTAGTCGAATG 2330
1589 .....AsnLeuAsnProAsnAl 1594
2331 CGATACACTTATACAGTCAGCAGCCACCCAAAAGGCAACCTTA 2380
1594 aglnValSerPheAsnGlnAlaAsn....PheAsnAsnAlaAsnValT 1609
2381 GCCTCGTGGGC..... 2391
1609 hrPheTyrGlyIleProLeuPheGlyLysThrProAsnPheGlyAsnSer 1625
2392 .....AATGCCCAAGCAACATTAAATCAAGCCACATT 2423
1626 ValArgLeuIleAsnPheLysGlyAspAlaLysPheAsnGlnAlaThrIle 1642
2424 AAMC..... 2427
1642 uAsnLeuArgAlaLysAsnIleHisLeuAsnPheGlnGlyAlaSerThrP 1659
2428 .....GGCAACNCAATCGGNTTGGGCAATGCTTCATT 2460
1659 heGluAsnAsnSerThrMetAsnLeuAlaGluSerSerGlnAlaSerPhe 1675
2461 AATCTA.....AGCAACAAGCCGCGCAAAAAGCGAGCT 2495
1676 AsnAlaLeuSerValGluGlyLthrAsnPheAsnLeuAsnGlySerSe 1692
2496 GACGCTTCCGACAAAGCT.....AAGCAAAAGTAAAGCCATT 2533
1692 rLeuLeuSerPheAsnGlyAsnSerValPheAsnAlaProValAsnPheT 1709
2534 CGCACTCAAGCAATGTCCTTACCGAATAGCAGATATTCATTT 2583
1709 yAlaAsnAsnSerGlnIleSerPheThrHisSerAlaThrPheAsnAla 1725
2584 GAAAACAGCCGCTTACCGCACTAGCGGAGCAAGAACAGCAT 2633
1726 AspAlaSerPheAspLeuGlyAsnAsnSerThrLeuAsnPheGlnSerVa 1742
2634 ACACCTTAAAGACAGCAATGACCTCCGCTCAGC..... 2670
1742 lLeuLeuSerAlaLeuAsnLeuLeuGlyAsnGlyLysAsnLeuA 1759
2671 .....ACGGAATTAGCAATTAACCTTGAACAAGCCACCATTAACATC 2715
1759 lAlaLeuAlaLysGlyAsnPheSerPheGlySerGlnGlyLe...Leu 1774
2716 AATCCGCTTATCGCCAGATGCTGACAGCGCAAAACGGCAGNGTTC 2765
1775 AsnLeuSerTyrMetAsnLeuPheGlyLysLysAlaSerValTyr 1791
2766 AGACAGCGCGCGCCGCTGCGCGCTCCCTA..... 2799
1791 rAspValLeuGlnAlaGlnAsnIleAspGlyLeuArgGlyAsnAsnGlyT 1808
2799 ..... 2799
1808 yTGlLysIleLeuPheTyrGlyIleGlnIleGlnLysAlaAspTyrSer 1824
2800 .....TTATCGCTTACACCGCCCACTTCGGTGA 2838
1825 PheAsnAsnGlyValHisSerTyrSerPheThrAsnPheLeuAsnThrTh 1841
2829 ATCCGCTTCAACAGCTGAGGTAAAGCAATTTGAACNGTCAA... 2874
1841 rGluThrIleThrGlnThrLeuHisAsnAsnArgLeuLysValGlnIleS 1858
2875 .....GGACATTCCGCTTATGTCGAACATC 2901
1858 erGlnAsnGlyAlaSerAsnAsnAlaMetPheAsnLeuAlaProSerLeu 1874

```

```

2902 TTCGCTACCGAGCAGCAAAATTGAAGCTGGCGAAAGTTC..... 2943
1875 TyrAspTyrGlnGlnAsnProTyrAspGluSerLysAsnSerTyrAsnH 1891
2944 .....GAAGNACTTACACCTTGGCGGTCACAAATACCGCA 2980
1891 sThrSerAspLysAlaGlyThrTyrTyrLeuSerSerIleLysGly. 1907
2981 ACGAACCCGTTAGCCTCGATCAATTGACGGTAGTGAAGGAAGAACAC 3030
1908 .....PheGlyLysAsnAsn 1912
3031 AAA...CCGCTGTCCGAAAACCTT...AATTTCACCTCGCAAAAACACA 3074
1913 GluIleProGlyThrTyrAsnAlaGlnAsnGlnProLeuGlnAlaLeuH 1929
3075 CGTCGATGCCGCGCGTGGCGTTACCAATTCATCCGCAAAAGCGGAGT 3124
1929 sile..... 1930
3125 TCCGCTGCATTAATCCGCTCAAGAACAGACCTTTC.....GACAAA 3168
1931 .....TyrAsnGlnAlaIleSerLysGlnAspLeuAsnMetIleAlaSer 1945
3169 CTCGGCAAGCGCAGCAAGCCAAAACAGCGGAAACACAAACGCGCAAG 3218
1946 LeuGlyLysGluPheLeuProLysValAlaLysLeuIleAlaSerGlyAl 1962
3219 CTTTGACGCGCTGATTTGGCGCGCGGATGCCCGCGGAAAGACAGCAA 3268
1962 aLeuAspAsnLeuAsnLeuAsnSerProAspSerPheGluThrIlePheS 1979
3269 GCGTT 3273
1979 erIle 1980

```

seq_name: p1r2:H98323

seq_documentation_block:

hypothetical protein AGR_L_3085 (imported) - Agrobacterium tumefaciens (strain C58, C:Species: Agrobacterium tumefaciens
C:date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:accession: H98323
R:Goodner, B.; Hinkle, G.; Gelling, S.; Miller, N.; Blanchard, M.; Ouello, B.; Gold
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
A.; Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:status: preliminary
A:molecule type: DNA
A:Residues: 1-1341 <KUR>
A:Cross-References: GB:AE007870; PIDN:AAK90114.1; PID:915160106; GSPDB:GN00170
A:Gene: AGR_L_3085
A:Map position: linear chromosome

alignment_scores:

Quality:	257.00	Length:	1373
Ratio:	0.421	Gaps:	67
Percent Similarity:	44.501	Percent Identity:	19.811

alignment_block:

US-09-303-518D-651 x H98323 ..

Align seg 1/1 to: H98323 from: 1 to: 1341

```

631 GCGGATTATCTATCCGCGCA...TGGTTAATTGGCGGCAATACACA 677
151 G1YThrAlaSerPheIleGlyAlaAsnTTPAlaAsnAspGlyAspAlaSe 167
678 TATGACAGGTTGGGAAATTAATGCGTANTTACTTTGACGCGGATGTC 727

```

```
167 rValIglYAsnGlyThrGlySerLeuLeuLeuHisGlyThrThra 184
172 GCGATGCGCAACGACATGAGCCATGCGATGCGATGCGGCGGCGC... 774
184 laserSerGluAsnIle.....TyrValIglYAsnSerGlyThr 196
775 .....GACAGCGGTGCGCAATGTTATTTATGCA 803
197 GlySerGlySerLeuLysLeuAspAsnSerSerThrLeuThrValAlaAs 213
804 CAAACA.....ACAAATTAATGCGTCTCAACAGGATTTTACAA 844
213 pThrValIleAlaIglYThrAsnSerThrPseSerSerAla..... 226
845 CCGGTACACCTTATTCGCGAGAAAGCGTTCCAGCTGATACGCAAA 894
227 .....AlaGlyGlySerGlyAsnValIglValIleuGlyAla 238
895 GATTGGTTCTAC.....GATGACATTTACAGAGCGGATACATACCGT 938
239 SerSerLeuYSerAlaAsnGlyValIleuAlaAsnAspProAspThrVa 255
939 CTMTTTGAACCGCGCATRACGACATTTTCTTACATCCACACAA 988
255 IglYThrAlaLeuValSerGlyGlySerSerThrPseMetSerSerA 272
989 AC.....GGTACGCGGTACGTAACAGAAACCAACAGAA 1020
272 spleuValValIglYValIglYValIglYThrLeuThr..... 284
1021 AAGGTTCCAAATCCAAAGCTTAAGTACAGACAGCTCGATGTTGACGA 1070
285 .....IleThrGlyGlyGlyMetValIleSerGlnThrThrValIleAl 300
1071 ATCTTTGAATGAACGTGATTAAGAACAGCTTTACGCGGCGAGGGGTGA 1120
300 aaSPLeuAsnSerAlaAspLysSerSerValThrValSer..... 313
1121 ATCAGTACCGTCCAGGTTTAAACAGCGTGAACCTTTCTTTATTCAT 1170
314 .....ArgLeuGlySerSerLeuGlnSerAspSerLeuVal 325
1171 TACGCAACGCGCAACTCATTTATCAACACATCAACAGCGCGGG 1220
326 ValIglYAsnGly.....GlyAlaAla 332
1221 CGGTTTATTTGAAGTGATTTTACGGTCTGCTGAAACACAGAAA 1270
332 alYsLeuValValIglValAlaAlaAlaThrValThr..... 343
1271 CGTGGCAAGC.....GCGGCGGTTCATATCAGTGAAGACACTTACC 1311
344 .....SerGlyGluAlaIleIleGlyAlaArgHisSerAlaSerGluAlaThr 358
1312 GTTACT.....TGG.....AAAGTAAACGG 1331
359 ValThrGlyAspGlySerLysTrpThrThrGlyAspLeuGlnValIglY 375
1332 CGTGGCAACGACGCGCTGTCAAAATCGGCAAGCAGCTGACGTTTC 1381
375 yaSPThrSerAspProGlyIglYLeuAlaGlyAsnGlyThrLeuAsnValT 392
1382 AAGCCAAAGGGAA.....AACCAAGGCTGACGACAGC 1413
392 hrAlaIglYIleSerValAspSerThrValAlaAlaHisLeuGlyValAla 408
1414 GTGCGGAGCGGTACAGTCAATTTGGATCAGCAGCAGAGATAAAGCAA 1463
409 GlyAlaThrGlySerAlaIleValAsp.....GlyIly 419
1464 AAAACAAGCCTTACT.....GAAATCGGCTTGNTCA 1495
```

```
419 sGlySerValTrpThrValAspArgAsnSerLeuGluValGlyVal...S 435
1496 GCGGCAAGGCGTACAGTCAACTGAATGACCGGTAATCAAGTCAACCCGAC 1545
435 erGlyAlaGlySerLeuAlaValThrGlyGlyGlyLeuValAspAlaAla 451
1546 AAACCTATTTTCGCTTCCGCGCGA..... 1572
452 AsnIleIleIleGlyThrAsnThrGlyGlyAsnGlySerValArgAlaSe 468
1573 .....CGTTGGATTTAAAGCGGATTCG 1597
468 rGlyAlaAspSerThrValLysSerArgSerAspLeuAsn...ValGlyL 484
1598 TTTCGTTCCACCGTATTCAAATACCGATGAAGGCGGATGATGNCAT 1647
484 eutyrglyAsnGlySerMetThrValGluAlaGlyGlyAlaValLysSer 500
1648 CATAT.....GCCACA.....ACAATCCACCGTTACCAT 1679
501 ArgAspGlyThrValAlaThrTyrGlyGlySerThrSerAlaValThrVa 517
1680 TACAGGGAATGAAGTATTACACAACGAGTGAAGATATCAATAGAC 1729
517 IThrGlyAspGlySer..... 522
1730 TTAATTCAGCAAGAAATTTGCTACAAAGGTTGTTGCGGAGAAAGAT 1779
523 .....SerTrp..... 524
1780 ACGACCAAAACGACGCGGCGCTACCTGTTTACGACCGCGCGGACGA 1829
525 .....AlaMetThrGlyThrPhePheValIglYThrLaserGlyAlaThr 539
1830 AGACCGACCCNCGCTTCCGCGGA.....ACAATTTAA 1867
539 rGlyAsnValThrValSerAsnGlyGlyAlaIleArgAlaThrGlyAlT 556
1868 AC...GGCAACATCAGCGCAACAAAGGCAAACTGTTTTCAGCGGC... 1911
556 hrLeuGlyAspLeuAlaGlyLaserGlyThrMetThrIleThrGlyAla 572
1912 .....AGCCGACACCGCCTACAT.....CATTTAAGGAAG 1946
573 GlySerLysValThrAlaThrValAlaAspAsnGlyThrValAsnSerGlySe 589
1947 C.....GGTGTCAAAATGAAGTATCCCAACAG 1978
589 rValAspValGlyPheGlnGlySerGlySerLeuSerValValAsnGlyG 606
1979 GAGAAATCGTGTGGACACAGCTGATCNACCGCAGTTTAAAGCGGA 2028
606 LysSerLeu.....Asp 609
2029 AATTTCATATTTCAGGCGGCGGCGTATTTCCCGCAATGTGGCAA 2078
610 AlatyraAsnLeuTyraGlyAsnAlaLeuGlySerSerGlyAlaValIle 626
2079 AGTGAAGCG...GATTGNCATTTGAGCAATCAGCCCAAGCAAGCAATTTT 2125
626 uValSerGlyValGlySerHisValSerValAspGlyLeuMetValValG 643
2126 GTGTGCAACCGCATCAAAAGCCATCAATCTGTACACGTTGCGACTGCACN 2175
643 LysAlaIglYAsnGlySerValGluIle.....Thr 653
2176 GGTCTGACAAATGTGTGAANAANCAATTACGACGATAAAGTATTC 2225
654 GlyGlyAlaSerLeuAlaAlaProThrIle.....LeuIleAl 666
2226 TTCATGTACATGAACNGACNTNAGCGCANCTNAGNCTNACNATNAGC 2275
666 a.....ThrGluAlaGlySerThrGlyValLeuSerIleIleValIglYs 681
```

```
2276 NTNNNTNAAACNTCNCNGGCTGNCNCACTNAAANGCAATCTTAGTGCA 2325
      ::::::::::::::::::::
681 ercIyglInthrAlaArgSerAlaGlyAlaValaGluAlaArgAlaIleAla 697
      ::::::::::::::::::::
2326 AATGGGATACAGCGTTATACAGTCAAGCCCAAGCCCAAGGCGC.. 2373
      |||
698 pheGlyAlaGlyAsnGlySerIleValaPheAsnHisSerGluThrGlyTyr 714
      ::::::::::::::::::::
2374 .....AACCTTAGCCTCGTGGCAATGCCCAAGCAACATTTA 2410
      ::::::::::::::::::::
714 rThrLeuSerAlaAspIleSerGlyAlaGlyAlaValaValaGluAlaG 731
      ::::::::::::::::::::
2411 ATCAAGCCCAATTAACGCGCAACNCATCGGNTTCGGCAATGCTTCATTT 2460
      ::::::::::::::::::::
731 LyAlaThrThrLeuSerGlyAsnAsnSerTyrSerGlyGlyThrThrIle 747
      ::::::::::::::::::::
2461 AAT.....CTAAGCAACAGCGCGCAAAAGCGCAAGTGTGACGT 2501
      ::::::::::::::::::::
748 SerAlaGlyMetLeuLysGlyThrAlaLysSerPheGlySerGlyGlyI 764
      ::::::::::::::::::::
2502 TTCGCGACGCTAAGGCAAGCAATGCGATTCGCGACACCAAGCGCATG 2551
      ::::::::::::::::::::
764 eValaAsnAsnAlaGluLeuValaValaAspGly.....GlyGlyThrL 778
      ::::::::::::::::::::
2552 TCTCCCTACCGCATAGGCGAGTATTCATTTTGA..... 2586
      ::::::::::::::::::::
778 euSerAsnAlaIleSerGlyThrGlySerPheGluLysThrGlyAspGly 794
      ::::::::::::::::::::
2587 .....ACAGCGCGCTTACCGGA..... 2604
      ::::::::::::::::::::
795 AsnLeuLeuThrGlyAsnSerThrTyrSerGlyAlaThrAlaValaSe 811
      ::::::::::::::::::::
2605 .....CAACTCAGCGCGAGCAAGANACAGATTAACACTTA.. 2640
      ::::::::::::::::::::
811 rAlaGlyLysLeuSerValaAsnGlySerLeuAlaSerAlaValaIle 828
      ::::::::::::::::::::
2641 .....AAA 2643
      ::::::::::::::::::::
828 LysSerGlyAlaThrValaGlyGlyThrGlyThrIleGlyGlyLeuThrVal 844
      ::::::::::::::::::::
2644 GACAGCGAATGACGCTGCCGTCAGCGACGAGATTAAGCAATTTA..AA 2690
      ::::::::::::::::::::
845 AsnSerGlyGlyThrLeuAlaProGlyAsnSerIleGlyThrLeuThrSe 861
      ::::::::::::::::::::
2691 CCTTGACAGCGCACCATTAACATTCATTCGCGCTAT.....GGCGACG 2734
      ::::::::::::::::::::
861 rThrGlyAsnAlaThrPheAlaSerGlySerThrTyrAlaValaGluIle 878
      ::::::::::::::::::::
2735 ATGCTGCGAGCG.....GGCGAAACCGCGCAGNGTGCAGAC 2769
      ::::::::::::::::::::
878 sPAlaAspLysSerSerAspArgLeuAlaValaThrGlyThrIleThr... 893
      ::::::::::::::::::::
2770 ACGCGCGCGCGCGTTCGCGCGCTCCCTATATCCGTTACACCG..... 2814
      ::::::::::::::::::::
894 .....IleAlaAsnAspValaSerLeuIleValaThrLeuGlu 906
      ::::::::::::::::::::
2815 .....CCAACTCGGTAGAAATCCGTTTCAACACGCTGACGGTAA 2854
      ::::::::::::::::::::
906 yAlaHisSerAlaTyrSerLeuGlyThrArgTyrThrIleLeuThrAla 923
      ::::::::::::::::::::
2855 ACGGCAATTCAGACNGTCAGAGACATTCGCTTATGTCGAACTCTTC 2904
      ::::::::::::::::::::
923 hrGlyGlyValaThr.....GlyThrPheSerSerValaAspLysSerPhe 937
      ::::::::::::::::::::
2905 GGCTACCGGACGCAAAATGGAAGTGGCGGAAGTCCGGAAGACTTA 2954
      ::::::::::::::::::::
938 AlaTyrLeuThrAlaLysValaIleGlnSerGlyAspAspAlaThrThr 954
      ::::::::::::::::::::
2955 C.....ACCTTGCGCGGTCA 2968
      ::::::::::::::::::::
954 rLeuSerPheLeuArgThrSerProAspSerGlyLeuLeuAlaIleAla 971
      ::::::::::::::::::::
```

```
2969 ACAAATACCGGACGCAACCCGTAAGCCATGATCAATTAGCGTAGTGAA 3018
      ::::::::::::::::::::
971 hrSerThrAlaAsnAlaArgAlaIleAlaAsnAlaValaGluAlaLeu... 986
      ::::::::::::::::::::
3019 GGGAAAGCAACAACACCGCTGTCCGAAAACCTTATTTCACCCCTGCAAAA 3068
      ::::::::::::::::::::
987 ...GlyGlnThrSerProLeuTyrGluAlaIleLeuThrLeuGlnGlnGln 1002
      ::::::::::::::::::::
3069 CGAACACGTCGATCCGGCGCGTGGCGCTTACCAACTCATCCGCAAGACG 3118
      ::::::::::::::::::::
1002 yGlu.....ThrGlnGlyAlaPheSer...GlnLeuAla.....G 1013
      ::::::::::::::::::::
3119 GCGAGTTCGCGCTCATTAATCCGCTCAAAAGCAAGAGCTTTCGACAAA 3168
      ::::::::::::::::::::
1013 LysGluIle.....HisProSerLeuAlaMetAlaIleuLeuAsnArg 1026
      ::::::::::::::::::::
3169 CTCGGGCAAGCGCAAGCCAAAAAACAGCGGAAAAAGACAAACGCGCAAG 3218
      ::::::::::::::::::::
1027 SerGly.....Glnse 1030
      ::::::::::::::::::::
3219 CCTTGAGCGCGTGTATTCGCGCGCGCGGCGATGCCCGCAAAAGACAGAA 3268
      ::::::::::::::::::::
1030 rArgAspValaIleLeuAsnArgLeuArgSerAlaPheGluGlyValaAsp. 1046
      ::::::::::::::::::::
3269 GCGTTGCCGAACCGCGCGCGCGAGCGGCGGAAAAATGTCGCGCATATG 3318
      ::::::::::::::::::::
1046 ..... 1046
      ::::::::::::::::::::
3319 CAGCGGAGAGAGAGAAAAAACGGGTGACAGCGGATTAAGACAGCGCNTT 3368
      ::::::::::::::::::::
1046 ..... 1046
      ::::::::::::::::::::
3369 GGGCAAAACAGCGGCAACCGGAACCCGCGCGGNTACACCGCCTTCCCC 3418
      ::::::::::::::::::::
1046 ..... 1046
      ::::::::::::::::::::
3419 GCGCCCGCNGCGCCCGCGGATTTGCGGCAACGCGACGCCCAACGCA 3468
      ::::::::::::::::::::
1047 .....AlaArgProIleLeuPro..... 1052
      ::::::::::::::::::::
3469 CCTCACCCCAACCGCAGCGGACCTGTATNAGCGCTTATGCCAATAGCGG 3518
      ::::::::::::::::::::
1053 .....ValaAlaTyrAlaGluGly 1059
      ::::::::::::::::::::
3519 TTTAGTGAATTTTCCGCCACACGCTCAACAGCGTTTCGCCGTACAGAGC 3568
      ::::::::::::::::::::
1059 yGlnAsnProPheAla..... 1064
      ::::::::::::::::::::
3569 AATTGACCGCGGTGTTGCCGAAGACCGCGCAACGCGNTTGGACAGC 3618
      ::::::::::::::::::::
1065 .....ValaAspAspLysAlaLeuSerPheTyrPseSer 1075
      ::::::::::::::::::::
3619 NGCATCCGAGNACCAAAACACTACCGTTGCGAAGATTCGCGCGCTACCG 3668
      ::::::::::::::::::::
1076 GlyPheArg..... 1078
      ::::::::::::::::::::
3669 CCAACAAACGACGCTCGGCCAAATCGGTATGCAAGAAAAACCTGCGACG 3718
      ::::::::::::::::::::
1079 .....SerArg 1081
      ::::::::::::::::::::
3719 GCGCGCTCGGCATCCGTTTTCGACAAACGCGACGCAAAACANCTTCGAC 3768
      ::::::::::::::::::::
1081 LysArgIleAsp.....Ser 1085
      ::::::::::::::::::::
3769 GAGCGCATTCGCAACTGCGACGCTTGCCACGCGCGCTT...TTCCG 3815
      ::::::::::::::::::::
1086 AspLysAsnGlySerSerValaAspMetAsnGlyGlyGlyLeuPheGln 1102
      ::::::::::::::::::::
3816 GCATATCGCATCGCGAGGTTTCGACATCGGCAATCAGACGCGCGGCGGT 3865
      ::::::::::::::::::::
1102 yLeuAspGly.....GluLeuSerAspSerTyrPArgAlaGly 1115
      ::::::::::::::::::::
3866 TTAGCAGCGGCGANTCTNTCAGACGGCATCGAGGCAAAATCCGCGCGCG 3915
      ::::::::::::::::::::
```

```

1115 .....||| ||||| |||
1116 aAlaAlaLeuTyrGlyHisAspGlyLeu.....LysGlnThrAlaLeu 1129
3916 GTGGCTCATTAAGCGCATTCAGGACAGCATATCCGCCGGTTTCGGGGGAAAT 3965
1130 AlaAlaSerAlaAspValAspSerTyrTyrLeuAlaAlaTyrAlaGly 1146
3966 CGGCATCGAAGCG.....TACATCGCGCAACGCGCTATTTCGTCC 4006
1146 rGlnIleGlyProAlaSerLeuArgLeuGlyAlaIleHisAlaPhe...G 1162
4007 AAAAAGCGATTACCGCTACGAAACGTCATATGCGCACACCCCGCTTT 4056
1162 InaSplAaGluThrArg.....ArgSerIle 1170
4057 GCGTTCACCGNATCCGCGGCGCATTAAGCAGATTAATTCATTCAAAC 4106
1171 SerPheSerThrLeuGlnGluSerLeuSerAlaSerTyrGlyGlySerTh 1187
4107 GCGCAACACATNTCC..... 4122
1187 rSerGlnValPheAlaGluAlaAlaThrParPheAspPheAspLeuThr 1204
4123 ..ATCACCCTTATTNAGCTGTCTCTATACCGATCGCGCTTCGGGCAAA 4170
1204 IsIleGluProTyrAlaAsnIleAlaTyr..... 1213
4171 GTCCGAACGCGGTCAAT.....ACCGNGTATTTGGCTCA 4205
1214 ValAsnThrArgThrAspGlyPheGlnGluLysGlyAlaIleAlaAla 1230
4206 GGATTTTCGCAAAACCGCG 4224
1230 IserSerGlySerThrArg 1236

```

seq_name: pir2:C82199

seq_documentation_block:

rxn toxin REXA VC1451 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)

C.Species: Vibrio cholerae

C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C.Accession: C82199
 R.Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
 Chaudson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.
 L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A.Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A.Reference number: A82035; MUID:20406833

A.Accession: C82199

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-4558 <HEI>

A.Cross-references: GB:AE004223; GB:AE003852; NID:99655942; PIDN:AF94608.1; GSPDB:GN001

A.Experimental source: serogroup O1; strain N16961; biotype El Tor

A.Gene: VC1451

A.Map position: 1

alignment_scores:
 Quality: 255.00 Length: 1750
 Ratio: 0.332 Gaps: 90
 Percent Similarity: 43.943 Percent Identity: 20.343

alignment_block:

US-09-303-518D-651 x C82199 ..

Align seg 1/1 to: C82199 from: 1 to: 4558

```

37 CGCAAGACCCGAAACCGCGCATCGCTTCCTCCCTGCTTACTTACG 86
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 ArgAsnAspProGlnThrGlyIleuLysTyr..... 375

```

```

87 CATATGCCCTGCTGGCATTTCTCCCAAGCTGGCGGACACACTT 136
376 .....TyrSerThrAlaTrp.....T 381
137 ATTTCGGCATCACTAC.....CATATCAATCGGCACTTTCGGCA 177
381 TyrLysGluValAsnHisLeuSerAsnLeuAlaAsnGlnAspIleSerAsp 397
178 AATAAGCGCAAGTTTGCAGTC.....GGGCG.....AAGA 209
398 AsnGlyGlyPheThrAlaValAsnIleAsnGlyAlaTyrThrLeuSerAs 414
210 TATTAGGCTNACACAAAAAGG.....GAGTTGCTGGCAAT 250
414 PleuLysValGluHisGlnSerValThrValHisAlaValGluLys 431
251 CAATGACAAAAAGCCCGATGATTTCTCTGCGTGTGCGCAACGC 300
431 erLeuThrGluTyrGluTrpAlaThrTyrAla.....AsnGly 443
301 GTGGCGCATTTGTGGCGCATATATATTTGTGACGCGCAGATACAG 350
444 ..AlaValIleAspAlaLysGluValSerLeuSerAspAlaLysMetG 459
351 CGGCTAT.....AACACGTTGATTTGGTGGCAG 382
459 YGlyHisAlaIleTyrAlaAspGlyThrLysValAspValLysAlaVal 476
383 GAAGNAATCCCGATCAGACCGCTTTCTTACCAATTTGCAAAAGAAAT 432
476 ySerAsnArgGlnProAsnThrTyrIleTyrAlaLysVal..... 489
433 AATTATAAGCCTGACAATTACACCTTACACGCGGCAATTANCATATGCC 482
490 .....LeuGlyProTyr..... 493
483 GCGTTGCATTAATTTGTACAGATGCAAGACCTGTGAATGACAGTG 532
494 .....ThrLysIleValIleValGluLeuAlaAsnA 504
533 ACATGAGGCGGATACCTATTCGATTAAGAAATAATTCGCGAGCTGTC 582
504 sProGluThrGlyAlaLeu.....LysTyrGlnAlaArgSer 516
583 CGCATCGCGTCAGACACACCTATTTGGCGTTAT...GATGATGCAAAACA 629
517 TrpTyrLysGluGlyAspHisThrAlaAsnIleAlaAsnGlnAspIle 533
630 CGCGCATTTATCTTACTCCGCGCATGTTAATTTGGCGCATATACATA 679
533 rSerAlaThrGlyTyrAsnPrometGlyLysGlyLysTyrSerLeuSerA 550
680 TGCAGGCTGGGAAATAATGCGCTANTTAGTTTGCAGCGGAGTGTGCG 729
550 spleuHisTyrSerValAsnAlaValArgSerThrSerGluThrValAla 566
730 CATGCCACGACTATGCGCCCTATGCGGATTCGCGGCGAGCGAG 779
567 AspIleGluGluTyrThrAspGlnThrLeuPheLysProAlaAsnAsp 583
780 CGGTTCCGCAATGTTTATTATGACAAACAAACAATAAATAGCTGTCA 829
583 rGly.....GluSerSerGlyAspValArgPhe 593
830 ACGAGGTTTACAAACGCGTACCTTATTCGCGAGGAAACGTTTC 879
593 snGlyAla.....GlyGly 598
880 CAGCTGATACCAAGATTTGTTCTACGATGACATTTACAGAGGCGATAC 929
599 AsnValIleLys.....SerAsnValThrArgGlyAsnVal 610
930 ACAT.....ACCGTCNTTTTGAACCGCGCAGTA 958

```



```
610 HispHeasnglyglyIleAlaasnValIleLeuHisserSergInp 627
959 AGGACATTTTCCCTTACATCCACACACAGCGTACGGTACGTAACA 1008
627 heglyAsnThrGluPhe.....AsnglyglyIleAlaAlaasn 639
1009 GAACCAAGAAAGTNTCCATCCAAAGCTTAAAGTACAGACAGTCG 1058
640 ValIleValIysSergIgluclIglYaspIleuThrPheargGlyAlaGl 656
1059 ACTGTTTACCAATCTTGAATGAAACTGAT.....AAGAACCACTTT 1102
656 yLeuAlaasnValIleuValHisGlnSergIgluclIglYasmetAspValT 673
1103 ACGCGGCGGGGGTGTAAATCAGTACCGTCCAAAGTTAAACACGGTGA 1152
673 yTAlaIglyIAlaValaasn...ValIleuValArgLeuGlyAspGlyGln 688
1153 AACCTTCT...TTATCATGATTACGGCAAC..... 1179
689 TyrIleuAlaHisLeuLeuAlaTyrGlyAsnIleSergValGlnYsGlySe 705
1180 .....GCCAACTCATCTTATCAACACATCAACCA..... 1212
705 rGlyAspSeraArgValaValmetLeuGlyGlyTyrFasnThrHisThrGlnI 722
1213 ..GGCGCGGGC...GGTTGTATTTT...GAAGGTGATTTACGGTCCG 1254
722 legIySergIYasnGlyLeuThrPheuAlaIaIglyIYpHeasnValmet 738
1255 CCTGAA..... 1262
739 ThrGlnValIglyIYsGlyAspValAlaIaValIleuAlaIglyIAlaAs 755
1263 CAAGCAACGCTGGCAAGCGCGCGC..... 1287
755 nValIleuThrYsmetGlyGluGlyIuLeuThrSergIYmetLeuGlyG 772
1288 .....GTTCATATCAGTGAAGAC.....AGTACGGTT 1314
772 yAlaAlaasnValIleThrHisIleSeraSnAspArgInLeuSeraSnThr 788
1315 ACTTGGAAGTAAACGGCGGTGCAACAGCCCTGTCCAAATCGGCA 1364
789 ThrIaValaIaIeuglyGlyAlaasn...IleuThrIYsIYsGly 804
1365 AGGC...ACGTGCACGTTCAAGCCAAAGGGAACCAAGCTCGATCA 1411
804 sGlyAsnThrLeuAlaValaValmetGlyGlyAlaasnVal...LeuThrH 820
1412 GCGTGGCGGCGGTACAGTC..... 1431
820 IsValIglyAspGlyThrThrThrGlyAlaMetValIglyIAlaasnIle 836
1431 ..... 1431
837 LeuThrIYsValIglyAsnGlyAspThrThrGlyIleLeuLeuGlyValGl 853
1432 ....ATTTGATCAGCAGCAGCAGTAAA.....GCCAAAAAACAG 1471
853 yAsnValIleuThrHisValIglyAspGlyGlnThrLeuGlyValaMetGly 870
1472 CCTTGAAGAATCGCTTGNTCAGCGGAGGGGTAGC.....GTG 1512
870 IaIaIglyAsnIlePheThrIYsValIglyAspGlyThrSeriIleAlaVal 886
1513 CAACGTAAATCCGATATCACTTCAACCCGACAAACTCTATTTCGCTT 1562
887 MetIleGlyIaIglyAsnIlePheThr.....HisValIglyI 899
1563 TCGGGCGGAGCTTGGATTAAACGGGCATTCCTTCTCCACCGTA 1612
899 uGlyAsnAlaTPrAlaLeuMetGlyGlyLeuGlyAsnValPheThrIYsV 916
1613 TTCAAAATACCAGTGAAGGGCG...ATGATTGNCNATCATTAATGCCACA 1659
916 aIglyAsnGlyAspAlaIleuAlaLeuMetValaIaIaIaValaPhe 932
1660 ACA.....ACATCCACCGCTTACCATTAACAGGAA 1688
933 ThrHisIleIYsPglYmetSerValaIaLeuMetLeuAlaYsGlyS 949
1689 TGAAGTATTATCACAAACCGAGTGTAAAGAT.....ATCAATA 1726
949 nValaIaIaThrIYsValIglyAsnGlyThrThrIleuAlaIaMetValGly 966
1727 GACTTAATTACAGCAAGAATTTGCTACACCGTGTGTTGGCGGAGAA 1776
966 snValaSnIlePheThrHisIleGlyHisGlySeraThrPhe..... 979
1777 GATACGACCAAAACGAAGCGGCGTCAACCTTGTTCACAGCGCGCGC 1826
980 .....AlaIaIaMetIleGlyGlnAlaAsnIleMetThrIYsValGly 994
1827 AGAAGACCGCACCCNGCTGCTTCCGGCGAACAAATTAACGGCACA 1876
994 nAspLeuThrAlaIaIaLeuMetValaIglyYsAlaAsnIleMetThrHisY 1011
1877 TCACGCAACAAACGGCAACGTTTTCACGCGCGACGACGACA..... 1920
1011 aIglyAspGlyThrSeraLeuGlyLeuPheAlaIglyIuValaSnValmet 1027
1921 .....CCGCA 1925
1028 ThrIYsValIglyAsnGlyThrThrIleuAlaIaIaMetPheGlyYsAlaAs 1044
1926 CGCCTACATCATTTAGGAAGCGGTGTCAAAATGGAAGTATGCCAC 1975
1044 nIleMetThrHisValaIglyAspGlyLeuThrGlyValaIleuAlaIeugly 1061
1976 AAGGAGAAATCGTGTGGACACAGCTGATCAGCGCAGCTTAAAGCG 2025
1061 IuAlaSnIleVal.....ThrIYsLeuGly 1069
2026 GAATTTTCCATATTCAGGCGGCGGCGGAGTATTCGCCGAATTTGC 2075
1070 AspAspPhe.....MetGlyValaIaIaIaIaIaIaIaIaSnVala 1084
2076 CAAGTGAAGCGGATTGN..... 2094
1084 IThrHisValaIYsAspAlaThrThrAlaAlaValIleuAlaIglyIY 1101
2094 ..... 2094
1101 snIleLeuThrIYsValIglyIuclYThrThrValaIYleuLeuIleSer 1117
2095 .....CATTTGAGCAATCAGCC.....CAAGCATTTTGGTGT 2129
1118 AspValaIYasnValaMetThrHisValaIglyAspGlyThrThrIleGlyI 1134
2130 CGCACCAGTCAAAAGCCATACAAATCTGTACAGCTTCGAC..... 2169
1134 eAlaIYsGlyIYsAlaAsnIleuIleThrIYsValaIYsPglYleuGlyY 1151
2170 .....TGGACNGGTCGACAAATTTGTCTGCAAAANCAATTACC 2208
1151 aIasnValIaThrTrp...GlyGlnAlaAsnValaPheThrGlnValaIYasp 1166
2209 GACGATAAAGTGATGCTTCA.....TTGACTTAAGC 2240
1167 GlyAspArgTyrAsnPheAlaIYsGlyIuAlaAsnIleIleThrIYsVa 1183
2241 NGACNTNAGCGGCANTGTNAGCNTNCCNATNAGCNTNTNTNAAACNTCN 2290
1183 IglyAspGlyGlnIuValaIserVal.....ValG 1193
```

```
2291 CNGGNTGNNACGTAAGCAATCTAGTCAATGAGTACACGT 2340
1193 lnglgluAlaasnlellethrhlsval...glyasnlyaspasp... 1207
2341 TATACA.....GTCAGCCACA 2357
1208 TyrThrGlyAlaTrpGlyLysAlaasnValleThrLysValGlyHisGly 1224
2358 C.....GCCACCCAAAACG 2371
1224 ygluAlaasnValleuAlaLysGlyLysAlaasnleValleThrGlnValG 1241
2372 GCAC.....CTTAGC 2382
1241 LysPglLysPserPheAsnAlaLeuTrpSerLysGlyAsnleValleThr 1257
2383 CTCGTGGCAAT.....GCCACGACACATTTAATCA 2414
1258 LysValGlyAspGlyMetGlnValleThrAlaLysGlyGlnAlaasnll 1274
2415 AGCCACATTAACGGCAAC.....NCATCGGNTTCGGCAATG 2452
1274 eThrThrThrValGlyAsnGlyLeuAsnValleThrAlaLysGlyAspa 1291
2453 CTCATTTAAT.....CTAAGCAACACGCCGCA... 2481
1291 laasnleAsnThrLysValGlyAspGlyValSerValAsnValleThr 1307
2481 ..... 2481
1308 GlyLysTyrAsnleAsnThrLysValGlyAspGlyLeuAsnValleAla 1324
2482 .....CAAAGCGAGTGTGAGCTTCGCAACGCTAAGG 2518
1324 lMetLysGlyLysAlaAsnAlaasnleHisValGlyAspGlyLeuAsnll 1341
2519 CAAAGCTTAACGATTCGCACTCAAC.....GGCAAT... 2550
1341 laasnleAsnThrLysValGlyAsnValleAlaLysValGlyAsnGly 1357
2551 .....GTCTCCCTAGCC.....GATTAAGC 2570
1358 AspPheTyrSerleuAlaValaLaserSerAsnThrSerSerAsnLysle 1374
2571 AGTATTCATTTGAAGACGCGCTTACCGCACTCAAGCGGACGA 2620
1374 uSerAlaLeuPheAspAsnllLysGlnThrValleuGlyValGlyGlys 1391
2621 AGGAACACGATTTACATTAAGAAGAGCAATGAGCGCTGCGTCAAGC 2670
1391 erGlnAlaLleAsnThrLysValGlnGlyAspGlyAlaLaserSerSerGly 1407
2671 AGGGAATTTAGCAATTTAACTTGACAAGCCACCATTTACCTCAATTC 2720
1408 ThrHisLysGlyArgGlyAlaLleAlaThrProGlnleThrLysLeuAs 1424
2721 CGCCTATCCGACGATGCT.....GCAGCG 2746
1424 pglYpHeGlnMetAspAlaLleLysGlnValSerSerAspLeuGlyAsps 1441
2747 GCCAAACGGCAGNGTGTCA.....GACACGCGCGCGCGCTTCGCGC 2790
1441 erleuThrGlySerValleThrLysValAspThrProAspLeuAsnLysMet 1457
2791 CGTTCCTTTATCGCTTACACCGCAACTGCGTAGATCCGTTGCA 2840
1458 GlnHisAlaLeuAsnVal...AspAspSerSerValGlnAla.....Pr 1471
2841 CACGCTGACGCTAAGCGCAAA..... 2862
1471 OAsnleuileValaasnGlyAspPheGlnleuGlyGlnHisGlyTrpGlns 1488

2863 .....TTGAC 2868
1488 erThrHisGlyValGlnAlaSerTyrAlaGlySerValleTrpGlyValGln 1504
2869 NGTCAGGAACATTCGCTTTATGTGGAACCTTGGCTACCGAAGCA 2918
1505 GlyGlnGlyHisGlyAlaArgValleThrGlnleuAspThrTyrThrAsnTh 1521
2919 CAATTC.....AAGTGGCGGAAGTTCGGAAGGACTTACACCTGG 2962
1521 rSerleuTyrGlnAspLeuAlaAsnleuAlaGlnGlyGlu...Vallele 1537
2963 CGGTC.....AACATACCGC 2979
1537 laValSerPheAspPheAlaLysArgAlaGlyLeuSerAsnAsnGlnGly 1553
2980 AACGACCCGTAAGCTCGATCAATTGACGCTGTGGAGGGAAGACAA 3029
1554 lleGlnValleuTrpAsnGlyGlnVallePheSerSerSerGlyAspGly 1570
3030 CAAACCGCTGCGCA...AACCTTAATTCACCTGCAAAACGACACG 3076
1570 uSerAlaTrpGlnGlnLysAsnleuLysleuThrAlaGln..... 1583
3077 TCGATCCGCGCGCGCTTACCACTCATC.....CGCAAAAC 3117
1584 .....AlaGlySerAsnArglleGlnPheLysGlyThrGlyHisAsnAsp 1598
3118 GGC..... 3120
1599 GlyleuGlyTyrilleuAspAsnValleAlaThrSerGlnSerSerGly 1615
3121 .....GAGTTCGCGCTGATTAATCCGCTAAGAAC 3151
1615 nglnAlaAsnAlaLleArgGlnHisAlaThrGlnAsnProAlaAlaGln 1632
3152 AAGAGCTTTCGCAAACTCGCAAGCGCAGAACCAAAAAACG...GCG 3198
1632 snAlaLeuSerAspLys...GluArgAlaGlnAlaAspArgGlnArgLeu 1647
3199 GAAAAAGCAACGCGCAAGCTTGACGCGCTGATTCGCG... 3237
1648 GlnGlnGlnLysGlnLysGlnleuAspAlaValleAlaGlySerGlnSerGly 1664
3238 .....GCCGCGC 3244
1664 nleuGlnSerThrAspGlnGlnAlaLeuGlnAsnAsnGlyGlnAlaGln 1681
3245 GCGATGCGCGC...GAAACACAGAAAGCTT...GCCAAGCGCGCGC 3288
1681 rGAspAlaValleLysGlnGlnSerGlnAlaValleThrAlaGlnleuAlaLys 1697
3289 CNGCGACGCGGGGAAATGTC..... 3309
1698 leuAlaGlnGlyLeuAspValleuAspGlyGlnAlaThrHisThrleGly 1714
3310 .....GCA 3313
1714 uSerGlyAspGlnTrpArgAsnAspPheAlaGlyGlyLeuLeuAspGly 1731
3314 TTATGCAAGCGGAGAGAAAGAAAAAGCGGTGACGCGGTAAGACAGC 3363
1731 alGlnSerGlnleuAspAspAlaLysGlnleuAlaAsnAspLysleAla 1747
3364 GCNTTGGCAAGACCGGAGGAAACCGCGGCTACACCGCTT 3413
1748 Ala...AlaLysGln.....ThrleuSerAspAsnSerle 1759
3414 CCCCCGCGCGCGCGCGCGCGGATTTGCGCAACCGACCCCAAC 3463
1759 sValLysGlnSerValAlaLysSerGlnAlaGlyValleAlaGlnGlyGlu 1775
3464 CGCACTCAACCCCAACCGCAGCGGACCTGATTCGCTTATGCAAT 3513
```



```
406 nglySerGlyAspIleAla.....LysThr.... 414
      :::::|||||
815 ATAAATGGCTGCTCAACGAGTTTACAAACGGCTACCCCTATATCCGGC 864
      :::::|||||
415 .....GlyGlyPheValGluThrSerGlyHisIleSerLysLeuSer 426
      :::::|||||
865 AGGGAACAGGTTTCCAGCTGATACGAAAGATGGTCTTACGAT..... 909
      :::::|||||
427 IleGluSerAsnAlaIleValLysThrLysGluThrPleuLeuAspProAs 443
      :::::|||||
910 .....GACATTTACAGAGCGCATACATACCGC 937
      :::::|||||
443 pAspValThrIleGluAlaGluAspProLeuArgAsnAsnThrGlyIleA 460
      :::::|||||
938 TCNTTTTACCGCGCAGTAAAGCATTTTCCATTACATCCAAACAC 987
      :::::|||||
460 snAspGluPheProThrGlyThrGlyIleValAsnSerAspProLysAsn 476
      :::::|||||
988 AACGGTACGGGTACGGTAAAGAAACAAAGAAAGTTCATCCAAATCCAA 1037
      :::::|||||
477 SerGluLeuLysThr...ThrLeuThrAsnThrThrIleSerAsnThrLe 492
      :::::|||||
1038 GCTTAAAGTACAGACAGTCCGACTGTTTACGATCTTTGAATGAACGTG 1087
      :::::|||||
492 uLysAsnAlaTrpThrMet.....AsnIleThrAlaSerA 504
      :::::|||||
1088 ATAAAGAACCACTTTACGGCGAGGGGCTTATATCAGTACGCTCCAAAG 1137
      :::::|||||
504 rGlySerLeuThrValAsnSerSer..... 511
      :::::|||||
1138 TTAAACACAGCGTGAACCTTTCTTTATCGATTACGGCAACGCAAACT 1187
      :::::|||||
512 IleAsnIleGlySerAsnSerHisLeuIleHisSerLysGly..... 526
      :::::|||||
1188 CATCTTATCAAAACATCAACCAAGCGCGCGCTTTGTTATTGTAAG 1237
      :::::|||||
527 .....GlnArgGlyGlyIleValGlnIleAspG 536
      :::::|||||
1238 GTGATTTTACGGTCTCGCTGCTGAAACAAAGAAACGTCGCAAGCGCGGC 1287
      :::::|||||
536 LysAspIleThrSerLysGlyAsnLeuThrIleTyrSerGlyLysTrp 552
      :::::|||||
1288 GTTCATATCATGTAAGACAGTACCGTTACTTGAAAGTAAACGGCGGTGC 1337
      :::::|||||
553 ValAspValHisLysAsnIleThrLeu..... 561
      :::::|||||
1338 AAACGACCGCCGTGCCAAATGCGCAACGCAAGCTGCACGTTCAAGCA 1387
      :::::|||||
561 ..... 561
      :::::|||||
1388 AAGGGGAAACCAAGCTCGATCAGCTGGCGAGCGTACAGTCATTTTG 1437
      :::::|||||
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe 575
      :::::|||||
1438 GATCAGCAGCGCAGCATTAAGCAAAACAAAGCAAGCTTTAGTAAGTCG 1487
      :::::|||||
576 ..GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.. 590
      :::::|||||
1488 CTTGATCAGCGCGCAGGGTACGCTGCAACTGATCCGATATCATCGTTCA 1537
      :::::|||||
591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys... 603
      :::::|||||
1538 ACCCGCAAACTCTATTTCGGCTTCGCGGCGAGCTTTGGATTAAAC 1587
      :::::|||||
604 .....AspPheArgAlaAsnAsnValSerLeuAsn 613
      :::::|||||
1588 GGGCATTCGCTTTCGTTCCACCGTATTAATAATACGATGAAGGGCGCAT 1637
      :::::|||||
614 GlyThrGlyLysGlyLeuAsnIleIleSerSerValAsnAsn..... 627
      :::::|||||
1638 GATTGNCATCATTAATGCCACAAACATCCACCGTTACCATTTACAGGGA 1687
      :::::|||||
```

```
628 ....LeuThrHisAsn.....LeuSerGlyThrIleAsnIleSerGlyAl 641
      :::::|||||
1688 ATGAAGATATTACACACCGAGTGTGTAG..... 1716
      :::::|||||
641 snIleThrIleAsnGlnThrThrArgLysAsnThrSerTyrTrpGlnThr 657
      :::::|||||
1717 .....AATATCAATGACTTAATTATACAGCAACAAAGAAAT 1748
      :::::|||||
658 SerHisAspSerHisTrpAsnValSerAlaLeuAsnLeu.....GluThr 672
      :::::|||||
1749 TGCCCTACACGCGTTGTTTGGCGAGAAAGATACGACCAAAACAGCGGC 1798
      :::::|||||
672 rGlyAlaAsnPheThrPheIleLysTyrIleSerSerAsnSerLysGly.. 688
      :::::|||||
1799 GCGTCACACCTGTTTACGACCGCGCGCAAGACCGCACCGCTGCTT 1848
      :::::|||||
689 ..LeuThrThrGlnTyrArgSerSerAla..... 697
      :::::|||||
1849 TCCGCGGCAACAAATTTAAACGGCAACATCATCCGCAAAACAGCGCAACT 1898
      :::::|||||
698 .....GlyValAsnPheAsnGly.....ValAsnGlyAsnMe 708
      :::::|||||
1899 GTTTTTC..... 1905
      :::::|||||
708 tSerPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProA 725
      :::::|||||
1906 .....AGCGGCAACCGCACCGCGACCGCTTC..... 1932
      :::::|||||
725 snGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsn 741
      :::::|||||
1933 .....AATCA 1937
      :::::|||||
742 IleThrAlaThrGlyGlyLysSerValPhePheAspIleTyrAlaAsnH 758
      :::::|||||
1938 TTATGGAAGCGGTGCTCAAAATGGAAGTATCCCAAGAAAGGAATCG 1987
      :::::|||||
758 sSerGlyArgGly.....AlaGluLeuLysMetSerGluLeuA 771
      :::::|||||
1988 TGTGCGCAACGAGCTGATCACCAGCGACGTTTAAAGCGGAAATTTTC... 2034
      :::::|||||
771 snIleSerAsn.....GlyAlaAsnPheThr 779
      :::::|||||
2035 .....CATATTCAGGGCGGCGAGCGGTGATTTCCCGCAATGTTGC 2075
      :::::|||||
780 LeuAsnSerHisValArgGlyAspAspAlaPhe..... 790
      :::::|||||
2076 CAATGTGAAGCGGATTTGNCATTTGAGCATTCACGCCCAAGCAGTTTTC 2125
      :::::|||||
791 ..LysIleAsnLysAspLeuThrIleAsn..... 799
      :::::|||||
2126 GTTTCGCGACCGCATCAAAAGCCATATCTGTACAGCTTGCGACTGACN 2175
      :::::|||||
800 .....AlaThrAsnSerAsnPhe... 805
      :::::|||||
2176 GGTCTGACAAATGTTGTCGAAANAANCATTCAGCAGATTAAGTATTC 2225
      :::::|||||
805 ..... 805
      :::::|||||
2226 TTCATTGACTAAGACNACNTNAGCGGCANTGTNAGCTNCCNATNAGG 2275
      :::::|||||
806 ..SerLeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaArgAsnA 822
      :::::|||||
2276 NTNNTTNAAANCCTCNCNCGGCGNTGCNNCACTNNAAGCAATCTTAGTCA 2325
      :::::|||||
822 lAlaIleAsnSerThrTyrAsnIleSerIleLeuGlyLysValThrLeu 838
      :::::|||||
2326 AATGGCATACAGCTTATACAGTCACGCAACAGCCACCAAAACGCAAA 2375
      :::::|||||
839 GlyGly.....GlnAsnSerSe 844
      :::::|||||
2376 CTTAGCCTTCGTGGCAATGCCAAGCAACATTTAATCAAGCC..... 2418
      :::::|||||
844 rSerSerIleThrGlyAsn.....IleThrIleGluLysAlaAlaAsnV 859
      :::::|||||
```


C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GH:BA000018; PID:913702612; PIDN:BA043752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447

alignment_scores:
Quality: 252.00 Length: 1552
Ratio: 0.348 Gaps: 51
Percent Similarity: 46.714 Percent Identity: 17.332

alignment_block:
US-09-303-518d-651 x F90073 ..

Align seg 1/1 to: F90073 from: 1 to: 2271

```
163 CGGAGATTGCGGAATAAAGCAAGTTGCGAGTGGGCGCAAGATAT 212
    ||| ..... ||| ||| ||| |||
    20 ArgLeuTyrLysSerGlyLysAsnTyrValLysSerGlyIleLysGlnI 36
213 TGAAGTTACACAAAGGAGAGTGGTCGCAATCAATGACAAAG 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    36 eGluMetPheLysIleMetGly.....L 44
263 CCCCAGATGATGATTTTCTGTGTGCGCGTACGGC..... 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    44 eupProPheIleSerHisSerLeuValSerGlnAspAsnGlnSerIleSer 60
301 .....GTGGCGCATGGTGGCGGATCA 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyAla 77
324 ATATATTGTAGCGTGGCAGATATAC..... 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    77 abeHerThrValAsnMetLeuHisAspGlnGlnAlaPheAlaIleSerAsp 94
349 .....GGCGGCTAT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    94 IapProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110
358 AACACAGTTGATTTGGTCGGAAGAAATCCCGATCAGCACCGTTT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    111 AsnSerThrThrIleGluAlaSerThrSerThrAlaAspSer..... 124
408 TTCTTACCAATTTGTCAAAAGAAATTAATTAAGCTGCAATTCACAC 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    125 .....ThrSerValThrLysAsnSerSerSerValGlnThrSer.... 137
458 CTTACACGGCGATTATCATATATCCGCTTGACATTAATTTGTCACAGAT 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    138 .....AsnSerAsp.....ThrValSer 143
508 GCAGAACTGTCGAATGACAGATGACATGAGGGGAATACCTATTCGA 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    144 SerGluLysSerGluLysValThrSerThrThrAsnSerThrSerAsnGln 160
558 TTAAGAAATATCCGAGCGTGTCCGATCGGCTCAGGACACACTTT 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    160 nGlnIleLysLeuThrSerThrSerGluSerThrSerSerLysAsnThr 177
608 GGCATTATGATGACAAACAGCGGATTAATCTTACTCCGGCGCATGG 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    177 hrSerSerSerAspThrLysSerValAlaSerThrSerSerThrGlnGln 193
658 TTAATTGGCGGCAATACACATATGACAGGCTTGGGGAATTAATGCCGTANT 707
```

```
194 ProIleAsnThrSerThrAsnGlnSerThrAlaSerAsnThrSerGln 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 TAGTTTGGCGCGCATGTGGCCATGCCAAC..... 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 nSerThrThrProSerSerValAsnLeuAsnLysThrSerThrThrSert 227
739 ..GACTATGGCCCTATGCCGATT..... 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 hrSerThrAlaProValLysLeuArgThrPheSerArgLeuAlaMetSer 243
760 .....GCAGGTGGCGGACGCGGT..... 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 ThrPheAlaSerAlaAlaThrThrThrAlaValThrAlaAsnThrIleThr 260
784 .....TCGCCAATGTTTA 796
260 rValAsnLysAspAsnLeuLysGlnTyrMetThrThrSerGlyAsnAla 277
797 TTTATGACAAA..... 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 hrTyrAspGlnSerThrGlyIleValThrLeuThrGlnAspAlaTyrSer 293
808 .....ACAAACATAATG 821
294 GlnLysGlyAlaIleThrIleuGlyThrArgIleAspSerAsnLysSer 310
822 GCTGCTACAGAGATTTTTACAAACCGGCTACCCCTTATTCGAGGAA 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 eHisPheSerGlyLysValAsnLeuGlyAsnLysTyrGlnGlyHisGly 327
872 ACGGTTTCCAGTGAATGCGCAAGATTTGTTTACGATGACATTACGA 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 snGly..... 328
922 GCGCATACATACCGTCTTTTGAACCGCGAGTACGAGATTTTTC 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 GlyAspGlyIleGlyPheAlaPheSerProGlyValLeuGlyGlnThrGln 345
972 CTTTACATCCAAC.....AACACGTTACGGGTA 1000
345 yLeuAsnGlyAlaAlaValGlyIleGlyGlyLeuSerAsnAlaPheGly 362
1001 CCGTAACGAACCAACCAAGAGTATTCACAAAGCTTAAGTACAG 1050
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 heLysLeuAspThrTyrHisAsnThrSerLysPro..... 373
1051 ACACTCCGACTGTTTACGATCTTGAATGAATGAATGAATGAACACAG 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 .....AsnSerAlaAlaLysAlaAsnAlaAspProse 384
1101 TTAAGCGGAGGGGGTGTATATCAGTACCGTCAAGTTAAACAGCGTG 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 rAsnValAlaGlyGly.....Gly 391
1151 AAAACCTTTCTTTATC.....GATTACGCAACGCGCAACTATC 1191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 IapPheGlyAlaPheValThrThrAspSerGlyValAlaThrThrTyr 407
1192 TTATCAACACATCAACAGCGCGGCGGCTTGTATTTGAAGTGA 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 ThrSerSerSerThrAlaAspAsnAlaAlaLysLeu..... 419
1242 TTTTACGCTGCGCTGAAACAGCAAGAGTGGGAGCGGCGGCTTC 1291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 ....AsnValGlnProThrAsnAsn...ThrPheGlnAspPheAspIle 434
1292 ATATCAGTGAAGACAGTACCGTT.....ACT 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
434 snTyrAsnGlyAspThrLysValMetThrValLysTyrAlaGlyGlnThr 450
1318 TGAAGATAACGGCGTGGCAACGACCGCTGTCCAAATATGCGCAAGG 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
451 TrpThrArgAsn.....IleSerAspTrpIleAlaLysSerGlyThrTh 465
1368 CACGCTGCACGTTCAAGCCAAAGGGAAACCAAGCGTCGATCAC... 1413
465 rsnPheSerLeuSerMetThrAlaSerThrGlyAlaThrAsnLeug 482
1414 .....GTGGCGCAGGTACAGTCATTTTGGATCAGCAGCAGACGATAA 1458
482 lnglnValglInPheGlyThrPhe..... 489
1459 GCGAAAAAACAAGCCCTTAGTGAATCGGCTGTGTCAGCGCAGGGTAC 1508
490 .....GluThrGlnSerAlaValThrGlnValArg...Ty 501
1509 GGTGCACTGAATGCCGATATCAGTCAACCCGACAACCTCTATTTCG 1558
501 rValaspValThrThrGlyLysAspIleleProProLysThrThrYrSerG 518
1559 GCTTTCGGCGCGGACGTTTGATTTAAAGCGGCAATTCGCTTCCTCAC 1608
518 Ty..... 518
1609 CGTATTCAAAATACGATGAAGGGCGATGTCNCNATCATATGCCAC 1658
519 .....AsnValAspGln..... 522
1659 AACCAATCCACCGTTACCATTAACAGGAATGAATATTAACACACCGA 1708
523 .....ValValThrIleaspAsnGlnInSerAlaLeuThrAlaL 536
1709 GTGGTAAGATATCAATAGACTTAATTAACGCAAAAGAAATGGCTACAC 1758
536 ySgIlyrAsnThrSerValAspSerSerTyAlaSerThrTyAsn 552
1759 GGTGTGTTGGCGAAGATACGACCA.....ACGACGG 1796
553 .....AspThrAsnLysThrValLysMetThrAsnAl 563
1797 GCGGCTCAACCTGTGTAC.....CAGCCCGCGCGAG 1828
563 aglyInSerValThrTyTrpPheThrAspValLysAlaLeuProThrValT 580
1829 AAGACCGCAGCCNGCTGCTTCGCCGCGACACAATTTAAAGGCAATC 1878
580 hrValglYasnGlnThrIleGlnValGlyLysThrMetAsnProValVal 596
1879 ACGCAAAACAAACGCAACTGTTTTCAGCGCAGACCGACCGCACCC 1928
597 LeuThrThrThrAsp..... 601
1929 CTACATCATTTAGAGAGCGGGGTGTCAAAAATGGAAGTATCCACAG 1978
602 .....AsnGlyThrGlyThrValThrAsnThrValThrGlyLeuProSerG 617
1979 GAGAATGTGTGGGACAACGACGTCGATCNCNCCGACGTTTAAAGCGGA 2028
617 Ty.....LeuSerTyAspSerAla 623
2029 AATTTCATATTACAGGCGCGGCGGTGATTTCCCGCATGTGGCAA 2078
624 ThrAsnSerIleIleGlyThrProThrLysIleGlyInSerThrValTh 640
2079 AGTGAAGGCGCATGNCATTTGAGCATCAGCCCAAGCAGCACTTTGGTG 2128
640 rValValSerThrAspGlnAlaAsnLysSerThrThrThrPheThrI 657
2129 TC..... 2130
657 leAsnValValAspThrThrAlaProThrValThrProIleGlyAspGln 673
2131 .....GCACGCGATCAAGCAGCATCAATCTGTACAGCTTC 2165
674 SerSerGlnValYrSerProIleSerProIleSerProIleSerThrGlnAs 690

2166 GGAC.....TGCACNGCTCTGACAATGTG 2191
690 pAsnSerGlyAsnAlaValThrAsnThrValThrGlyLeuProSerGlyL 707
2192 TCGAANAAACATTAACGAGATTAAGTG..... 2220
707 eutThrPheAspSerThrAsnAsnThrIleSerGlyThrProThrAsnIle 723
2221 .....ATTGCTCATTTAGTACAGACGACGACNTNAGCGGANTGTNAG 2261
724 GlyThrSerThrIleSerIleValSerThrAspAlaSerGly..... 737
2262 NCTNNCCNATACGNTNNTNNAANCTCNCGGCGCTGCNNCACTNNAANG 2311
737 ..... 737
2312 GCAATCTTAGTCAATGGCGATACAGTATATACAGTACGACCAACGGC 2361
738 .....AsnLysThrThrThrThrPheLysTyGlnVal 748
2362 ACCCAAAAGCGC.....AACCTTAGCCTGTGGGCAATGCCAA... 2400
749 ThrArgAsnSerMetSerAspSerValSerThrSerGlySerThrGlnG 765
2401 .....GCAACATTTATCAAGCCACATTTAAAGCGACACNCAT 2437
765 nSerGlnSerValSerThrSerLysAlaAspSerGlnSerAlaSerThS 782
2438 CGGNTTGGGGCAATGCTTCATTAATCAAGCAAAACGCGGACAAAC 2487
782 erThrSerGlySerIleValValSerThrSerAlaSerThrSerLysSer 798
2488 GGCAGTGTGACGCTTTCGCAACGCTTAAGGCAAAAGTAAGCATTCGCC 2537
799 ThrSerValSerLeuSerAspSerValSerAlaSerLysSerLeuSerTh 815
2538 ACTCAACGGCAATGTCCTCAGCGGATTAAGGAGATTCATTTGAA 2587
815 rSerGlnSerAsnSerValSerSerSerThrThrSerLeuValAsnS 832
2588 ACAGCCGCTTTACCGCAACACTCAGCGGCGGCAAGGAMNACGATTAAC 2637
832 erGlnSerValSerSerSerMetSerGlySerValSerLysSerThrSer 848
2638 TTTAAGACAGCGAATGAGCGCTGCGTCAGGACGCAATTAAGCAATTT 2687
849 LeuSerAspSerIleSerAsnSerAsnSerThrGlnLysSerGlnSerLe 865
2688 AAACCTTGACACAGCCACCATTAACACTCAATTCGCGCTATCGCCACAGATG 2737
865 uSerThrSerThrSerAspSerLeuArgThrSerThrSerLeuSerAspS 882
2738 CTCGAGCGCGCAACCC...GGCAGNGTGTACAGACGCGCGCGCGCT 2784
882 erLeuSerMetSerThrSerGlySerLeuSerLysSerGlnSerLeuSer 898
2785 TCGCGCGCTTCCTATTATTCGCTTACACCGCCCACTTCGTTAGAACCCG 2834
899 ThrSerIleSerGlySerSerSerThrSerAlaSerLeuSerAspSerTh 915
2835 TTTCAACACGCTGACGTAACGCGCAAAATTTGAACNGTCAAGAACATFTCC 2884
915 rSerAsnAlaIleSerThrSerThrSerLeuSerGlnSerAlaSer... 930
2885 GCTTTATGTGGAACCTTCGCTACCGAAGGACGCAAAATTTGAAGCTGGCG 2934
931 .....ThrSerAspSerIleSerIleSer 938
2935 GAAAGTTCGGAAGNACTTACACCTTGGGCGCAACATATACGGGACAGA 2984
939 AsnSerIleAlaAsnSerGlnSerAlaSerThrSerLysSerAspSerG 955
```

```

2985 ACCGTAAGCTCATGATTTGAGGTAGTGGAAGGAAAGCAACAAC 3034
      : : : : : : : : : : : : : : : : : : : : : :
955 nserThSerIleSerLeuSerThr.....SerAspSerIys 968
3035 CGCTG.....TCGGAACCTTATTTACCCCTGCAAAAGCAACAGTC 3078
      : : : : : : : : : : : : : : : : : : : : : :
968 ermSerThSerIleSerLeuSerAspSerThSerThSerIleSer 984
3079 GATGCGCGCGGTGGCTTACCACATCCGCAAAAGCGCGAGTCCG 3128
      : : : : : : : : : : : : : : : : : : : : : :
985 ValSerGlySerLeu..... 989
3129 CCTGCATATATCCGTCGCAAAAGCAAGCTTCCGACAACCTCGCAAG 3178
      : : : : : : : : : : : : : : : : : : : : : :
990 .....SerIleAlaSerGlnSerValSerThSerThSerAsp 1004
3179 CAGAAGCCCAAAACAGCGCGGAAAGCAACGCGCAAGCGTTGACCG 3228
      : : : : : : : : : : : : : : : : : : : : : :
1004 ermSerThSerIleValSerAspSerIleSerThSerGlySer 1020
3229 CTGATTGGCGCGCGGCGATGCCGCAAAAGCAAGCGTTGACCG 3278
      : : : : : : : : : : : : : : : : : : : : : :
1021 LeuSerAlaSerAspSerIleSerMetSerValSerSerMetSerTh 1037
3279 ACCGGCGCGCGCGGCGGGAATGTCGCATATGACAGCGGAGG 3328
      : : : : : : : : : : : : : : : : : : : : : :
1037 rSerGlnSerGlySerThSerGlnSerLeuSerAspSerGlnSer 1052
3329 AAGAGAAAAACGGGTGACGCGGATTAAGACAGCGCTTGGCGAAG 3378
      : : : : : : : : : : : : : : : : : : : : : :
1053 .....ThSerAspSerAspSer.....LysSer 1060
3379 CGCGAAGCGGAACCGCGCGGTTACCAACCGCTCCCGCGCGCGCG 3428
      : : : : : : : : : : : : : : : : : : : : : :
1061 LeuSerLeuSerThr.SerGlnSerGlySerThSerThSerThSer 1077
1077 hrSerAlaSerVal.....ArgThSerGlnSerGlnSerThSerGly 1091
3429 CGCGCGCGCGGATTTCCGCAACGCGACCCCAACCGCAACCTCAAC 3478
      : : : : : : : : : : : : : : : : : : : : : :
1077 hrSerAlaSerVal.....ArgThSerGlnSerGlnSerThSerGly 1091
3479 AACCGCAGCGCGACCTGATNACCCTTATCCCAATAGCGGTTGAGTAA 3528
      : : : : : : : : : : : : : : : : : : : : : :
1092 SerMetSerAlaSerGlnSerAspSerMetSerIle..... 1103
3529 TTTTCCGCGACGCTCAACAGCGCTTTCCGCGTACAGCAATGAGCG 3578
      : : : : : : : : : : : : : : : : : : : : : :
1104 .....SerThSerPheSerAspSerThSerAspSerIys 1116
3579 CGTGTTCGCGAAGACCGCGCAACGCTTGGACAAGCNGCATCCG 3628
      : : : : : : : : : : : : : : : : : : : : : :
1116 erAlaSerThrAlaSerSerGlnSerIleSerGlnSerAlaSerThSer 1132
3629 AACCCAAACATACGCTTCCGAAGTTCCGCGCTACCGCGCAACAAAC 3678
      : : : : : : : : : : : : : : : : : : : : : :
1133 ThSerGlySerValSerThSerThSerLeuSerThSerAsnSerGln 1149
3679 GACCTTCGCGCAATCGGTATCGCAAAAACTCGGACAGCGCGCTCG 3728
      : : : : : : : : : : : : : : : : : : : : : :
1149 uArgThSerThSerValSerAspSerThSerLeuSerThSerGln 1166
3729 CATCTCTTTT.....CGCACACCGCGACGCAAAACATCTGACAGC 3772
      : : : : : : : : : : : : : : : : : : : : : :
1166 erAspSerIleSerGlnSerThSerThSerAspSerIleSerGlnAla 1182
3773 GCATTCGCAACTCGGACGCTTCCGCGCGCGCTTTCGGGCAATAC 3822
      : : : : : : : : : : : : : : : : : : : : : :
1183 IleSerAlaSerGlnSerThSerIleSerLeuSerGlnSerAsnSerTh 1199
3823 GGCATGCGAGGTTCGACATCGCATCAGCAGCGCGGCTTTAGACG 3872
      : : : : : : : : : : : : : : : : : : : : : :
1199 rSerAspSerGlnSerGlnSerAlaSerAla.....PheLeuSerG 1213
3873 CGGCANTCTNTCAGACGCGCATCGAGCAAAATCCGCGCGCTGCTGC 3922

```

```

1213 lserLeuSerGlnSerThSerThSerThSerIleSerValSerSer 1229
      : : : : : : : : : : : : : : : : : : : : : :
3923 ATTAGCGCATTCAGCGCATACCGCGCGGTTCGCGGATTCGCGATC 3972
      : : : : : : : : : : : : : : : : : : : : : :
1230 SerThSerIleSerThSerIleSerAspSerThSerIleSerIle 1246
3973 GAACCGTACATGCGCGCACCGCGCTATTGCTCCAAAAGCGATTACG 4022
      : : : : : : : : : : : : : : : : : : : : : :
1246 rThSerThSerIleSerAsnSerThSerGlySerAlaSerIle...S 1262
4023 CTACGAAAGTCATATGCGCCACCGCGGTCTGCTTCACCGATAC 4072
      : : : : : : : : : : : : : : : : : : : : : :
1262 erThSerThSerIleSer.....GlnSerThSerThr 1273
4073 GNCGCGGC.....ATTAGCGCATATTCATTCGCAACCGCGCAC 4115
      : : : : : : : : : : : : : : : : : : : : : :
1274 PheIleSerIleSerValSerThSerIleSerMetSerThSerTh 1290
4116 CATATTCATCACMCCTTATTNAGCTGTCTATACCGATCCGCTTCG 4165
      : : : : : : : : : : : : : : : : : : : : : :
1290 rLeuSerAsnSerThSerLeuSerThSerLeuSerAspSerThSer 1307
4166 GCAAAGTCCGACACCGCTCATATCCGCGGTATGCTCAGAT...TTC 4212
      : : : : : : : : : : : : : : : : : : : : : :
1307 sPheIleSerAspSerLeuSerThSerMetSerThSerAspSerIle 1323
4213 GCGCAAAACCGCGCATGCGGATGCGGTAACCGCGCAATCAAGTTT 4262
      : : : : : : : : : : : : : : : : : : : : : :
1324 SerThSerIleSerAspSerIleSerThSerThSerIleSerGly 1340
4263 CACG 4266
      : : : : : : : : : : : : : : : : : : : : : :
1340 rThr 1341

```

seq_name: p1r2:AF0351

seq_documentation_block:

probable autotransporter protein yapa [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AF0351

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.

deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0351

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1430 <KUP>

A:Cross-references: GB:AL590842; PIDN:CA92137.1; PID:q15980853; GSPDB:GN00175

C:genetics:

A:Gene: yapa

alignment_scores:

Quality: 250.50 Length: 1190

Ratio: 0.442 Gaps: 55

Percent Similarity: 47.647 Percent Identity: 20.000

alignment_block:

US-09-303-518D-651 x AF0351 ..

Align seg 1/1 to: AF0351 from: 1 to: 1430

```

955 AGTAAGGACATTTTCTTACATCCACAACAGCGTACGGTACGCT 1004
      : : : : : : : : : : : : : : : : : : : : : :
402 AladlglgIupheasprThralAlaIysasn.....AlaIadglute 415
1005 AACGAAACCAACGAAAGGTTCCATTCGAAGCTTAAGTCAAGCAG 1054
      : : : : : : : : : : : : : : : : : : : : : :

```


415 uAlaGluThrAlaAlaLysAla.....IleGluAla 426
1055 TCAGACTGTTTGACGAATCTTGAATGAAGATGAATGAACAGGTTTC 1104
426 IAlaYsIleThrAspLysAlaValGluAspAlaThrAlaAlaTyrLysGlu 442
1105 GCGGCGGGGGTGTATATCATGACGTCACGCTCCAGGTTTAAACAGCGTGAAA 1154
443 AlAlaAlaAspLysAlaGluGlnThrLysThrAlaLeuGluAlaGlu 459
1155 CCTTCTTTATCGATTACGCAACGGCAACTCATCTTATCAACACACA 1204
459 sAlaLys.....GluAspAlaAspLysLeuValValThrAsnThr 472
1205 TCACCAAGCGCGGGCTTGTATTTGAAGTGATTTACG..... 1248
473GlyLeuLeuAsnAspAlaAspGlnAlaLeuGlu 483
1249GTCTGCTGAAAACAAAC.....GAACGTGGCAAGCGCGG 1286
484 GlnLeuValThrAlaGlnAsnAsnAlaGlnProThrLeuAspLeuProAl 500
1287 CCGTATATCATGTAAGACAGTACCGTTACTTGAAGAAAGGCGCGTGG 1336
500 aIleAspValThrIleAlaProAlaLysThrGlnAspValIleGluGlyT 517
1337 CAACGACCGCCTGTCCAAATTCGGCAAGGCGACGCTGACGTTCAAGCC 1386
517 hSerAlaIleAlaThrGlnValAlaGlyGlyThrGlnAsnValAlaLys 533
1387 AAAGGGAACCAACAGCTCGATCAGCGTGGCGACGCTGACATGCTTTT 1436
534 GlyGlyLysAlaIleAspSerValIleThrLysAspGlyIleValAsnLe 550
1437 GGATCAGACGCGACGATTAAGCAAAAACACGCTTAGTGAAACG 1486
550 uAlaAlaGlyAlaAsnAlaLysGlyThrGlnValThr..... 562
1487 GCTTNTCAGCGCGGAGGTACGCTGACATGATCCGATATACAGTTC 1536
563LysGlyThr.....LeuAsnAsnAsnGlyGlyAl 572
1537 AACCCGCAAACTCTATTTGGCTTTCGGCGGACGCTTGGATTAA 1586
573 AspThrAspThrValVal.....SerThrGluGlyLysLeuValLeuTh 587
1587 CGGCGATTTCGCTTCCACCGTATTCAAAATACCGATGAAGGCGGA 1636
587 rGlyGlySer.....GluThrAlaIleAlaThrSerThrGlyAla 601
1637 TGATTGNCNATCATATATGCACACA..... 1662
601 ySvalAlaGluGlyGlyValValThrAlaGlyAspHisSerValIleGlu 617
1663ACATCCACCGTTACCTTACAGGGAATGAAGTATTAC 1700
618 LysMetIleSerSerGlyAsnValThrAlaSerGlyAsnAsnThrIleVal 634
1701 ACAACCGAGTGGTAAGATATCATATAGACTTAATTAC..... 1737
634 IArgAspThrThrIleAsnAspGlyLysLeuSerLeuAlaGlyThrAla 651
1738 ..AGCAAGAAATTCCTACACGCTGTTT.....GGCGAG 1773
651 hTrAlaAsnAsnThrThrPheAsnGlyGlyIlePheSerValGluGlyAsp 667
1774 AAAGTACAGCAACAAAGAAC.....GGGCGCTCAACCTTGTTTACA 1817
668 ThrAlaAlaThrLysThrAsnMetThrGlyLysPheAlaValThrG 684
1818 GCCCGCGCAGAAAGACGACCGCAGCTGCTTCCGGCGGAACAATTTA 1866
684 yAsnAlaLysIleGluGluThrValValSerAlaGlyAspValSerLeuA 701
1867AACGGCAATCACGCAACAAACGCAACTGTT 1902
701 IAspLysAlaThrAlaAsnAsnThrThrLeuThrGlyGlyThrPheAla 717
1903 TTCAGCGGCAACGACACCGCAGCCTACATATATTAGAACGGGTG 1952
718 AlAlaGlyAspThrAlaValSerAlaThrAsnMetThrGlyGlyThrPh 734
1953 GTCAAAAATGAGAGTATCCCAAGAGAAATGTGTGGGACACGACT 2002
734 eThr..... 735
2003 GGATCNCACGACGTTTAAAGCGGAATTTCCATATTCAGGCGGCGAG 2052
736 ..ValAlaGlyAspThrAlaAlaThrLysThrLysMetThrGlyGlyGlu 751
2053 GCGGTGATTTCCGCAATGTTGCCAAGTGAA..... 2085
752 PheAlaValThrGlyAsn...AlaLysIleGluAspThrValLeuAsnAl 767
2086 ..GGCGATTGNCATTTGAGCAATCACGCCCAAGAGTTTGTGTGCGCAC 2134
767 aSerAspPheSerLeuAlaAspLysAlaThrAla..... 778
2135 CGCATCAAGGCATACATCTGTACAGCTTCGACCTGACAGNGGTGACA 2184
779AsnAsnThrThrLeuThrAspGlyThrPheThr..... 789
2185 AATTGTGCGAANAANACATTTACCGACGATTAAGTGATTCCTCATGTAC 2234
790ValAlaGlyAsp.....AlaAlaValThr 797
2235 TAACACGACNTNAGCGCANTGNTAGCTNCCNATNAGTNNNTNAA 2284
797 rAlaThrAsnMetSerGly..... 803
2285 ANCTCNCNGGCGNTGNCNACTNAAAGCAATCTTAGTGAATGGCGAT 2334
804GlyLysPheAlaValLysGlyLys 811
2335 ACACGTTATACAGTACGCCACCAACGCCCAAAAGCGCAACCTTGACCT 2384
812 AlAlaLys.....IleLysAspThrGlnLeuSerAlaGlyAsnPheThrLe 826
2385 CGTGGCAATGCCCAACCAATTTATCAAGCCACATTTAAACGGCAACN 2434
826 uAlaGluAsn.....AlaThrAlaAsnAspThrThrLeuAsnGly.... 839
2435 CATCGGNTTCGGCAATGCTTCAATTAATCTAGCAACAACGCGGCA... 2481
840GlyLysPheAspValSerAsnGluAlaThrAla 850
2482CAAAACGCACTGACGCTTTCGACACGCTAA 2516
851 ThrAsnThrThrIleAsnAsnGlyLeuPheThrLeuLysAspGlyAlaH 867
2517 GCGAAAGTACGCAATTCGCACTCAACGCAATGCTCTCCATAGCGGATA 2566
867 sAlaAsp...SerThrThrValAsnSerGlyThrPheValMetAlaAspG 883
2567 AGCGAGTATTCATTTGAANAACAGCGGCTTACCGGACACACGCGG 2616
883 In..... 883
2617 AGCAAGANACACATTAACCTTAAAGACAGAGTAAGAGCGTGGCGTC 2666
884 SerThrAlaAsnGlyIleGlnLeuValAspSerAlaPheThrLeuAlaSe 900
2667 AGCGACGGAATTTAGCAATTTAAACCTTGACAAACGACCATTTACATCA 2716
900 rGlyAlaLysAlaSerGly.....IleThrLysL 910


```

1933 .....AATCATTTAGACGGGTGTGCTCAAAA 1960
530 pvalvalaspleuadlaaglyasnaspmethtisileglyaspjyl 547
1961 TGCAGGATAC.....CCACAAGAGAAATC 1986
547 ysaaspjlyvaltyrvalvaliileaspalaaglyaspjylvalserleu 563
1987 GTGTGGGACAACGACGTGATCMACCGACGTTTAAACGGGAAATTTCCA 2036
564 Alasnaspasnnglntyrleuglythrthr.....GI 574
2037 TATTCAGGGGGGCGACGGGTGATTTCCCAATGTTGCCAAGTGGAG 2086
574 nilealaserglythrleuemetvalseraspasn.....SerGlnleug 589
2087 GCGATTTGCATTTGACATCACCAGCCAGAGTTTGTGTGCGACCG 2136
589 lytyrthrhtstyrasnarg.....Glnvaliilepethrhaplypro 603
2137 CATCAAGC.....CATACATCTGTACAGTGTGGAGCTG 2171
604 GlnGlnSerValmetGlnIlethrAlaasnValaspthrArgSerThrth 620
2172 GACNGGTGACAAATGTGTGCAANAANCATTACGACGATTAAGTGA 2221
620 rthGlnhtsaglyaspjileglumetargalaspjylgluvalalav 637
2222 TTGCTTCATTAAGTAAACN.....GACNTNAGC 2250
637 alaspalaglyvalaspthrGlntrpGlyAlaIleuMetAlaAspSer 653
2251 GGCATGTGNAAGCTNNCCNATNAGNTNNTNAAANTCNCNCGGCMTC 2300
654 Glycin..HisGlnaspjlyserthrleuthrlysthrGlyAlaG1 669
2301 NNCATNNAAGCAATCTTAGTGCAAATGCGATACAGTTATACGTCA 2350
669 ythrleu.....GlnleuthrAlaserGlyThrThrGlnSerAlaVal. 683
2351 GGCACACGCCACCCAAAACGCAACTTACGCTGCGGCAATGCCAA 2400
684 .....ArgValGlnGlnGlnThrleuGln.....GlyaspValAla 695
2401 GCAACATTTAATCAAGCCACATTA.....AACGCAACNCATC 2438
696 AspilepneptotyralaserSerleuthrvalGlyaspjylAlaThrPh 712
2439 GGNITGGGCAATGCTTCATTTAATCTAAGCAACAACGCCGCAAAAACG 2488
712 evalthrGlyAlaspglnaspIleGlnSerIleaspAlaThrSerSerG 729
2489 GCAGTCTGACGCTTCGCAACGCTAAGGCAACGTAAGCATTC... 2535
729 lyThrIleaspIleSeraspjylThrValIleuArgleuthrGlyGlnasp 745
2536 .....GCACTCACGCGCAATGTCTCCCTAGCCGTAAGCGATAT 2576
746 ThrSerValAlaIleuasnAlaserleupheasnCyaspjylThrLeuVa 762
2577 CCAT...TTTGAACAACGCGCTTACCGGACACTCAGCGGCGACGAAG 2623
762 lasnAlaThraspjylValThrleuthrGlyGlnleuasnThrAsnleug 779
2624 ANACACATTAACCTTAAAGACGAGATGACGCTCCGTCAGGACG 2673
779 luthr.....AspSerleuthrThrleuSerAsnValThr 790
2674 GAATTAGCAATTTAAACCTTGACAACGCCACCATTAACATCAATTCGC 2723
791 ValasnGlyAsnleuthrAsnThrSerGlyAlaValaserleu..... 804
2724 CTATGCCACGATGTGTGACGGCGCAAAACGCGACAGNTGTCAACACGC 2773

```

```

805 .....GlnasnGlyValAlaaglyasp..... 811
2774 CGCGCGCGCTTCCGCGCTTCCTATTATCCGTTACACCGCCAACTTCG 2823
811 ..... 811
2824 GTAGATCCCGTTTCAACAGCGTGACGGTAAACGCAATTGACNGTCA 2873
812 .....ThrleuthrValasnGlyaspjylThrGlyG1 822
2874 AGCAACATTCGCTTATGTGCAACTCTTCCG...TACGACCG 2917
822 yGlyThrleuIleuAspSerGlnleuasnGlyaspSerValSerA 839
2918 ACNAATTTAGCTGGCGGAAAGTTCGGAAGNACTTACACCTTGGCGT 2967
839 spGlnleuValmetasnGlyasnThrAlaGlyAsnThrThrValAlaVal 855
2968 AACAA...ACCGC...AACGACCCGTAACCTCGATCAATTAGCGT 3011
856 AsnSerIleThrGlyIleGlyIuprothrSerThrGlyIleValValA 872
3012 AGTGAGGGAAGACACAAACCGCTGCGGAAACCTTAATTTCACCC 3061
872 lasphealAlaaspProthrGlnpheGlnAsnAlaGlnpHeSerL 889
3062 TGCAAAACGAA...CACGTGATGCCGCGCGTGGCGTTACCAACTCATC 3108
889 euAlaGlySerGlyTyralasmetGlyAlaItyrAspThrleuVal 905
3109 CGCAAGACGGCGAGTTCGCGCTGCATTAATCCGTCACAAACAGACCT 3158
906 GluAspAsnAsnAspThrleu..... 913
3159 TTCCGACAAACTCGGCAAGCAGAAACCAAAACAGGGGAAAGACA 3208
913 ..... 913
3209 ACGCGAAAGCTTGACGGCTGATTCGCGCGCGCGCGATGCCCGAA 3258
913 ..... 913
3259 AAGACAGAAAGCTTCCCGAACCGCGCGCGCGCGGAGGGAATATCT 3308
914 ArgSerGlnIleValThrProProser..... 922
3309 CGGCATTATCAGCGGAGAGAGAAAGAAACGCGTGCAGCGGATTAAG 3358
923 .....ProProAspProA 927
3359 ACAGCGCNTTGGGAAACAGCGGGAAGCGGAACCGCGCGNTACCAAC 3408
927 spProthr..... 929
3409 GCCTTCCCGCGCGCGCGCGCGCGCGGATTTCCGCAACCGAGCC 3458
930 .....ProAspProAspPr 934
3459 CCAACCGCAACCTCAACCCCAACCGACGCGACCTGATNAGCGTTATG 3508
934 othrGlnaspProAspProthrProAspProGlnProthrProAlaTyrG 951
3509 CCAATAGCGGTTGAGTGAATTTTCCGCAACGCTCAACAGCGTTTCCGC 3558
951 lnpValIleuAsnAlaIysValGlyGlyThrleuAsnAsnleuArgAla 967
3559 GTACAGACGAATTTGACCGCGCTGTTGCCGAGAGACCGCGCAAC... 3603
968 Ala.....AsnGlnAlaPheMetGlnIupArgIrgAspHisAl 960
3604 .....GCNGTTGGACAACGNCATCCGCAACCAAC 3637

```

```

980 aglyglaspiglyglnthrlleuasnleuargvallelgllyasptryh 997
3638 ACTACCGTTCGCAAGATTTCCGGCCCTACCGCAACAAACGACCTGGCC 3687
997 1stytthralalaelglleuaha.....glnhsluaspthrser 1011
3668 CAATCGGTATGCAAAAAACCTCGGACGGGGCC..... 3723
1012 ThrValGlnleuSerGlyasphepserglargtrpGlyThrAspGly 1028
3724 .....GTCCGATCCTGTTTCCGACAAACCGGACCAAAACNCT 3763
1028 yglutpmetleuuglyleval.....glyglut 1038
3764 TCGACGAGCGCATCGCAACTCGGACGCGTTCGCCGCGCGCTTTTC 3813
1038 yTserAspasnleuaspserargsermetthGlyThrArgala 1054
3814 GGGCAATACGGCATCGGACGTTTCGACATCGGATCAGACGGGCGCGG 3863
1055 AspaSnGlnAsnHslGly...TyrAlaValGlyleu...ThrSerSerTr 1069
3864 TTTTACGACGGGCANTCTTTCAGACGCGATCGGA..... 3897
1069 pHeGlnHslGlysglnlysglnGlyAlaTrpLeuAspAsnTrpLeug 1086
3898 .....GGCAAAATCCGGCGCGCGCTGCTG 3921
1086 lntTyrAlaTrpHeserAsnAspValSerGlnHsluaspGlyAlaAsp 1102
3922 CATTAC.....GGCATTCAGCAGATACCGCGCGCTTTC..... 3957
1103 HlStyHslSerSerGlylletlealaserleuAlaGlylTyGlnTr 1119
3958 .....GGCGATTCGGCATGCAACCGTACATCGGCGCACCGCCT 3997
1119 pLeuProGlyArgGlyValAlaIlleGlyProGlnAlaGlnValIlleTyg 1136
3998 ATTTCGTCGCAAAAGCGATTCACGCTACGAAAC..... 4032
1136 lnglyAlaGlnleuAspAspHethrAlaAlaAsnArgAlaArgValSer 1152
4033 .....GTCAATATCGCACACCCCGGCTTTCGCTTCACCCGNTA 4070
1153 GlnserGlnGlyAspAspIlleGlnTr..... 1161
4071 CGNCGCGGATTAAGCAGATTAATTCATCAACCGCGCAACACATNT 4120
1162 ArgleuGlyleuHslSerGlyTrp....ArgThrAlaValHsls... 1174
4121 CCATACNCCATTATTNAGCCTGTCC...TATACGATGCGCGCTTCGGGC 4167
1175 ..ValIleProThrleuAspLeuAsnTyTyrHslAspProHslSerThr 1190
4168 AAAGTCCGCAACAGCGTCATACCGCNCNTATTGGCTACAGATTTGGCA 4217
1191 GlnIlleGlnGluAspAlaSerThr.....IleSerAspAspAlaVally 1205
4218 AACCCGC...AGTGGGGAATGGGCGGTAAACCGCAATC...AAAGTT 4261
1205 sGlnArgGlyGlnIlleTyValGlyAlaThrGlnHslSerGlnArgv 1222
4262 TCACGCTTCNCTCCAGCTGCGCGCGCAAGNCCGCAACTGGAGAG 4311
1222 alSerleuArgGlySerValAlaTrpGlnGlySerAspAspHea 1238
4312 CAACACAGCGCGGATCAAAATTAAGCTACCGCTG 4347
1239 GlnThrAlaGlyPheleuSerMetThrVallystr 1250
seq_name: p1r2:F85862
seq_documentation_block:

```

```

hypothetical protein yfal [Imported] - Escherichia coli (strain 0157:H7, substrain E)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85862
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanita, E.; Potamousis, K.; Apod
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85862
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1250 <STO>
A:Cross-references: GB:AE005174; NID:g12516568; PIDN:AA657362.1; GSPDB:GN00145; UMG
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: yfal

```

alignment_scores:

```

Quality: 250.00      Length: 1412
Ratio: 0.399         Gaps: 66
Percent Similarity: 44.405      Percent Identity: 19.263

```

alignment_block:

```

US-09-303-518D-651 x F85862

```

```

Align seg 1/1 to: F85862 from: 1 to: 1250

```

```

646 TCCGCGCGCATGTTAATTCGCGCAATACATATGCAAGGCTTGGGAAA 695
95 ThrGlyMetThrleuPheAlaAsnAsnThrValSerGlylTyGlnAsn 111
696 TAAATGGCGTA.....N 706
111 nGlyGlyAlaIllePheAlaTyGlnAsnSerThrleuAsnleuThrAspY 128
707 TTAGTTGACGCGGATGTCGCCATGCAACGATATGACGCTATGCCG 756
128 alIlePheSerGlyAsnVal.....AlaGlyGlyTyGlnGlyAlaIlle 142
757 ATTGCAGGTGGCGGACGACGACGCGGTTCGCAATGTTATTATGCA 806
143 TySerSerGlyThrAsnAspThrGlyAla..... 152
807 AACAAACAATTAATGCTGCTCAACGAGATTTCACAACCGGCTACCTT 856
153 ....IleSpLeuArgValThrAsnAlaValPheArgAsnAla 168
857 ATTCCGCGCAGGAAACGATTTCAGCTGATACGCAAAAGT...TGGTTC 903
168 snAspGlyTyGlnGlyAlaIlleTyThrIleAsnAsnAspIleTyLeu 184
904 TAGCATGACATTTCACAGAGCGGATACACATACGCTCTNTTTTGA 953
185 SerAspAspValPhe.....As 190
954 CAGTAACGAGACATTTTCCTTACATCCAAACAACGAGTACGGGT... 999
190 nAsnAsnGlnAlaTyThrSerThrSerTyTyrSerAspGlyAspGlyAla 207
1000 .....ACGTAACAGAAACCAACGAAAGGTTNCCATTCACAAGCTTAA 1044
207 laIleAspValThrAspAsnAsnSerSerTyShSpro.....Ser 221
1045 GTACAGACAGTCGACGTGTTGACGAATCTTTGAATGAACCTGATAAGA 1094
222 GlyTyThrThrle..... 225
1095 ACCAGTTTACCGCGGAGGGGTGTTAATCAGTACCGTCAAGGTTAACA 1144
226 .....IleAsnA 228

```

```

1145 ACGGTAAACCTTTCTTTATGATTACGGCAAGCAACTCATCTTA 1194
1195 TCACACACATCAACCAAGCGGGGGTGTATTTTGAAGCGATTT 1244
232 ThrAsnThrAlaGluGlyTyrGlyAlaIleTyrThrAsnSerAl 248
1245 TACG.....GTCTGCGCTGAAAAACAAGCAAGCGTGC 1276
248 arThrAlaProTyrLeuIleAspIleSerAlaAspSerTyrSer...G 264
1277 AAGCGCGCGCGCTTCATCATGTAAGACAGTACCGTTACTTGGAAAGTA 1326
264 IAsnMetGlyValLeuValAspGluAsnAsnSerAlaIleGlyTyrGly 280
1327 AACCGCGGTGCAACGACCGCGCTGTCACAAATGCGCAAGCAAGCGCTGCA 1376
281 AspGlyPro.....SerSerAlaIleGlyIlePheMetCty 292
1377 CGTTCAAGCCAAAGGGGAAACCAAGCGTGCATCAGCGTGGCGGACGTA 1426
292 rLeu.....GlyLeuSerGlyValThrPheAspIleAlaAspGlyL 306
1427 CAGTCATTTTGCATCAGCAGCAGCAGATAAAGCAAAAACAGACCTTT 1476
306 yThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerIle 322
1477 AGTGAATACGGCTG.....NTCAGCGGCGAGGGTACGGTCAACTGAA 1520
323 AlaGlyThrGlyLeuIleThrIleThrGlySerGlyAspLeuValLeuAs 339
1521 TGCCGATTAAT..... 1530
339 nAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluV 356
1530 ..... 1530
356 alThrLeuGlyIleArgSerAsnSerLeuMetAsnValGlyAspThrHisCys 372
1531 CAGTTCAACCGCGCACAACTCATTTGCGCTTTCGCGGCGGCGGCTTGA 1580
373 GluAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerIleAs 388
1581 T.....TTAAAC..... 1587
388 pLysTyrGluAsnGlnAlaGluLeuAsnValGlySerThrGlnGlnThrP 405
1588 ..GGGCATTTCGCTTCGTTCCACGCT.....ATTCAAAATACGCGATGA 1629
405 heAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAspAla 421
1630 GGGGCGATGATGNCNATCATTAATGCCAACAAACATCCACCGTTACAT 1679
422 GlyGlyAsnValThrValAsnMetGlySerPheAlaGlyThrIleGluG 438
1680 TACAGGAGATGAAGATATTACACACCGAGTGTGAAGATATCATATAGAC 1729
438 yAlaGlyGlnLeuThrIleAlaGln..AsnGlySerTyrVal.....L 452
1730 TTAATTAACGCAAGAAATTTGCTTACACAGGTGGTTGGCGAGAGAGAT 1779
452 euAlaGlyAlaGlnSerMetAlaLeuThrGly.....Asp 463
1780 ACGACCAAAAGAAAGCGGCGCTCAACTGTTTACACGCCCGCGCGCA.. 1827
464 IleValValAspAlaGlyAlaValLeuSerLeuGluGlyAspAlaAlaAs 480
1828 .....GAAGACCGCACCCNCGTGCCTCCGCGGCA.. 1857
480 pLeuAlaIleLeuGlnAspAspProGlnSerIleValLeuAsnGlyGlyM 497
1858 ..ACAAATTTAAACGCGCAACATCACGCAAAACGCGCAAACTGTTTTC 1905

```

```

497 etLeuAspLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLys 513
1906 AGCGGACAGCCGACACCGCACCGCTAC..... 1932
514 AspGlyLeuGluValSerGlySerSerGlyThrValIleGlySerGlnAs 530
1933 .....ANTCATTTAGCAAGCGGGTGTGCANAAA 1960
530 pValValAspLeuAlaGlyGlnAsnAspMetHisIleGlyGlyAspGlyL 547
1961 TGAAGGTATC.....CCACAAGAGAAATC 1986
547 yAspGlyValTyrValValIleAspAlaIleGlyAspGlyGlnValSerLeu 563
1987 GTGTGGCAACAAGCAACGTCGATTCNCCGACGCTTTAAAGCGGAAATTTCCA 2036
564 AlaAsnAspAsnGlnTyrLeuGlyTyrThr.....G 574
2037 TATTCAGGCGCGCGGCGGATGATTCGCCCAATGTTGCCAAAGTGAAG 2086
574 nIleAlaSerGlyThrLeuMetValSerAspAsn.....SerGlnLeuG 589
2087 GCGATTGNCATTTGAGCAATCAGCCGCCAAGCAGTTTGTGGTGCACCG 2136
589 IyTyrThrHisTyrAsnArg.....GlnValIlePheThrAspLysPro 603
2137 CATCAAGC.....CATCAATCTGTACAGTCTCGGACCTG 2171
604 GlnGluSerValMetGluIleThrAlaAsnValAspThrArgSerThrTh 620
2172 GACNGGTGTGACAAATGTGTGCAANAAACATTTACGAGCAATAAGTGA 2221
620 rThrGluHisGlyArgAspIleGluMetArgAlaAspGlyGluValAlaV 637
2222 TTGCTTCATTGACTAAGACN.....GACNTNAGC 2250
637 alAspAlaGlyValAspThrGlnTrpGlyAlaLeuMetAlaAspSerSer 653
2251 GGCANTGTNAGNCTNNCCNATNAGNTNNTNNAACNTCNCNGGCGTGC 2300
654 GlyGln...HisGlnAspGluGlySerThrLeuThrIleThrIleGlyAlaG 669
2301 NNCATNAAAGCAACATCTGTGCAAAATGGCGATACAGTTATACAGTCA 2350
669 yThrLeu.....GluLeuThrAlaSerGlyThrTrpGlnSerAlaVal. 683
2351 GCCACAAGCGCCACCACCAAAAGCGACCTTAGCCTGTGGCAATGCCCAA 2400
684 .....ArgValGluGluGlyThrLeuGln.....GlyAspValAla 695
2401 GCAAATTTAATCAAGCCACATTA.....AACGGCAACCAATC 2438
696 AspIlePheProTyrAlaSerSerLeuTyrPvalGlyAspGlyAlaThrPh 712
2439 GGNATCGGCAATGCTTCATTTAATCTAAGCAACACCGCGGCAACAAAGC 2488
712 eValThrGlyAlaAspGlnAspIleGlnSerIleAspAlaIleThrSerSerG 729
2489 GCAGTCTGACGCTTTCGACACACGCTAAGGCAAAACGTAAAGCATTC... 2535
729 IyThrIleAspIleSerAspGlyThrValLeuArgLeuThrGlyGlnAsp 745
2536 .....GCACTCAACGGCAATGTCTCCCTGACCGGATTAAGCAATATT 2576
746 ThrSerValAlaLeuAsnAlaSerLeuPheAsnCysAspGlyThrLeuVa 762
2577 CCAT...TTTGAACACAGCGCTTACCGGACAACTCAGCGGAGCAAG 2623
762 IAsnAlaIleThrAspGlyValThrLeuThrGlyGluLeuAsnThrAsnLeuG 779
2624 ANACAGCATTTACCTTAAAGACAGCGGAATGACGCTGCGGTACGACAG 2673

```


4312 CAACACAGCGCGGCATCAATTAGGCTACCGCTGG 4347
||| ::::: ::::: ::::: |||
1239 GlnThrAlaGlyPheLeuSerMetThrValLysTrp 1250
